Charles LANG, George SIEMENS, Alyssa WISE, and Dragan GAŠEVIĆ (Eds)

Handbook of Learning Analytics

FIRST EDITION

SOLAR
SOCIETY for LEARNING ANALYTICS RESEARCH

DOI: 10.18608/hla17

Editors:
Charles Lang
George Siemens
Alyssa Wise
Dragan Gašević

www.solarresearch.com

This handbook is dedicated to Erik Duval, in memoriam.

Erik's passion and strong convictions were some of the fundamental pillars on which the learning analytics community was built. His participation in this community and his scientific contribution to the field made him one of our leading experts. His personality and openness made him one of our favorite colleagues. His integrity and never-ending curiosity made him an example to follow.

All the editors and authors of this handbook were inspired by his openly shared ideas and research. By similarly sharing ours, we want to honor his memory, keeping the fire he helped start burning strong.

Introductory Remarks

At the Learning Analytics and Knowledge Conference in 2014, The *Handbook of Learning Analytics* was conceived to support a scientific view of the rapidly growing ecosystem surrounding educational data. A need was identified for a Handbook that could codify ideas and practices within the burgeoning fields of learning analytics and educational data mining, as well as a reference that could serve to increase the coherence of the work being done and broaden its impact by making key ideas easily accessible. Through discussions over 2014-2015 these dual purposes were augmented by the desire to produce a general resource that could be used in teaching the content at a Masters level. It is these core demands that have shaped the publication of the Handbook: rigorous summary of the current state of these fields, broad audience appeal, and formatted with educational contexts in mind.

To balance rigor, quality, open access and breadth of appeal the Handbook was devised to be an introduction to the current state of research in the field. Due to the pace of work being done it is unlikely that any document could capture the dynamic nature of the domains involved. Therefore, instead of attempting to be a terminal publication, the Handbook is a snapshot of the field in 2016, with the full intention of future editions revising these topics as research continues to evolve. We are committed to publishing all future editions of the Handbook as an open resource to provide access to as broad an audience as possible.

The Handbook features a range of prominent authors across the fields of learning analytics and educational data mining who have contributed pieces that reflect the sum of work within their area of expertise at this point in time. These chapters have been peer reviewed by committed members of these fields and are being published with the endorsement of both the Society for Learning Analytics Research and the International Society for Educational Data Mining.

The Handbook is composed of four sections, each dealing with a different perspective on learning analytics: Foundational Concepts, Techniques and Approaches, Applications and Institutional Strategies. The initial section, *Foundational Concepts*, takes a broad look at high-level concepts and serves as an entry point for people unfamiliar with the domain. The first chapter, by Simon Knight and Simon Buckingham Shum, discusses the importance of theory generation in learning analytics, Ulrich Hoppe examines the history and application of computational methods to learning analytics, and Yoav Bergner provides a primer on educational measurement and psychometrics for learning analytics. The final chapter in the section is Paul Prinsloo and Sharon Slade's consideration of how the ethical implications of learning analytics research and practice is evolving.

The second section of the Handbook, *Techniques and Approaches*, discusses pertinent methodologies and their development within the field. This section begins with a review of predictive modeling by Chris Brooks and Craig Thompson. This introduction to prediction is expanded upon by Ran Liu and Kenneth Koedinger in their discussion of the difference between predictive and explanatory models. The next chapters deal with differing data sources, the emerging sub-field of content analytics is covered by Vitomir Kovanović, Srećko Joksimović, Dragan Gašević, Marek Hatala and George Siemens. The theme of unstructured data is continued with an introduction and review of Natural Language Processing in learning analytics by Danielle McNamara, Laura Allen, Scott Crossley, Mihai Dascalu and Cecile A. Perret and further augmented by an expansive discussion of related methods in discourse analytics by Carolyn Rosé. Sydney D'Mello then delves into the strides that emotional learning analytics have made both within the learning and computational sciences. From this exploration of the depths of the inner world of students Xavier Ochoa's chapter on multimodal learning analytics covers the expansive range of ways that student data may be tracked and combined in the physical world. In the final chapter in the *Techniques and Approaches* section, Alyssa Wise and Jovita Vytasek discuss the processes involved in taking up and using analytic tools.

The third section of the Handbook, *Applications*, discusses the many and varied ways that methodologies can be applied within learning analytics. This is followed by a consideration of the student perspective through an introduction to how data-driven student feedback can positively impact student performance by Abelardo Pardo, Oleksandra Poquet, Roberto Martínez-Maldonado and Shane Dawson. This is followed by a thorough introduction to the uses, audiences and goals of analytic dashboards by Joris Klerkx, Katrien Verbert, and Erik Duval. The application of a theory-based approach to learning analytics is presented by David Shaffer and Andrew Ruis through a worked example of epistemic network analysis. Networks are further explored through

Daniel Suthers' description of the TRACES system for hierarchical models of sociotechnical network data. Applications within the realm of Big Data are tackled by three chapters, Peter Foltz and Mark Rosenstein consider large scale writing assessment, René Kizilcec and Christopher Brooks look at randomized experiments within the MOOC context and Steven Tang, Joshua Peterson and Zachary Pardos consider predictive modelling using highly granular action data. Automated prediction is further developed in a tutorial on recommender systems by Soude Fazeli, Hendrik Drachsler and Peter Sloep. This is followed by a thorough introduction to self-regulated learning utilizing learning analytics by Phil Winne. After which Negin Mirriahi and Lorenzo Vigentini consider the complexities of analyzing user video use. The final chapter in the *Applications* section considers adult learners through Allison Littlejohn's analysis of professional learning analytics.

The final section of the Handbook, *Institutional Strategies & Systems Perspectives*, is designed to help readers understand the practical challenges of implementing learning analytics at an institutional or system level. Three chapters discuss aspects pertinent to realizing a mature learning analytics ecosystem. Ruth Crick tackles the varied processes involved in developing a virtual learning infrastructure, Linda Baer and Donald Norris speak to challenges of utilizing learning analytics within institutional settings and Rita Kop, Helene Fournier and Guillaume Durand take a critical stance on the validity of educational data mining and learning analytics for measuring and claiming results in educational and learning settings. Elana Zeide provides an in-depth analysis of the state of student privacy in a data rich world. The final chapters in the Handbook discuss linked data, Davide Taibi and Stefan Dietze cover the utilization of the LAK dataset and Amal Zouaq, Jelena Jovanović, Srecko Joksimović and Dragan Gašević consider the potential of linked data generally to improve education.

Ultimately, our aim for the Handbook is to increase access to the field of learning analytics and enliven the discussion of its many areas and sub-domains and so, in some small way, we can contribute to the improvement of education systems and aid student learning.

Charles Lang

George Siemens

From the President of the Society for Learning Analytics Research

As a field of academic study, Learning Analytics has grown at a remarkably rapid pace. The first Learning Analytics and Knowledge (LAK) conference, was held in Banff in 2011. The call for this conference generated 38 submissions and 130 people attended the meeting. By the next year, the conference had more than doubled the number of paper submissions, and registration closed early when the conference sold out at 230 participants. Both paper submissions and attendance numbers at LAK have increased every year since. Now in 2017, the 7th LAK conference returns to Vancouver with a program representing work from 32 countries, featuring 64 papers, 67 posters and demos, and 16 workshops. At the writing of this preface, the conference is expected to reach or exceed the 426 participants who attended LAK 2016.

The growth of learning analytics is not only demonstrated by a bigger and broader conference; the field has matured as well. In 2013 the Society for Research on Learning Analytics (SoLAR) was officially incorporated as a professional society, and the first issue of the *Journal of Learning Analytics* appeared in May, 2014. Publications of learning analytics research are not limited to the society's conference proceedings and journal. Special issues focused on learning analytics have appeared in journals concerned with education, psychology, computing and the social sciences. A Google Scholar search of the term "learning analytics" produces over 14,000 hits, coming from a broad range of journals such as *Educational Technology & Society*, *American Behavioral Scientist*, *Computers in Human Behavior*, *Computers & Education*, *Teaching in Higher Education*, and many others.

All of this shows that the body of research representing work in the learning analytics field has also grown rapidly. So much so that it is an opportune time to produce a learning analytics handbook that can serve as a recognized resource for students, instructors, administrators, practitioners and industry leaders. This volume provides an introduction to the breadth of learning analytics research, as well as a framework for organizing ideas in terms of foundational concepts, techniques and approaches, applications, and institutional strategies and systems perspectives. Given the pace of the research, it is likely that a static handbook would be out of date quickly. Therefore, SoLAR offers this volume as an open resource with plans to update the content and release new editions regularly. Meanwhile, this volume provides an extensive view into what we know now from the perspective of leading experts in the field. Many of these authors have been involved in learning analytics since that first meeting in Banff, while others have helped broaden and enrich the field by bringing in new theory, concepts, and methodologies. Further, readers will find chapters not only on technical and conceptual topics, but ones that thoughtfully address a number of important but sensitive issues arising from research utilizing educational data, including ethics of data use, student privacy, and institutional challenges resulting from local and national learning analytics initiatives.

As the newly elected president of the Society for Learning Analytics Research it is my great pleasure to welcome you to this handbook. My own history with the society dates back to the first Vancouver meeting where I found a fascinating group of researchers who shared my excitement about the questions we could now ask about learning by leveraging the increasing volumes of data produced by new educational technologies. These were data geeks for sure, but I also found people with a sense of purpose and strong commitment to diversity and inclusion. These values are reflected in the research we do and in the community we have built to advance that work. This volume reflects those values by its nature: the chapters come from authors around the world representing very different disciplinary training, whose work has an orientation to action and supporting educational advances for all learners.

With the publication of this volume we welcome old friends and introduce the field to newcomers. I am confident that you will find here research that will simultaneously excite you with ideas about how we can change the world of education and frustrate you with the evidence of the challenges ahead for doing so. We invite you to join us in this adventure. If you would like to become directly involved in learning analytics, SoLAR has several ways for you to do so:

- Join our professional society (SoLAR):
    - https://solaresearch.org/membership/

- Submit your work and/or attend our annual conference (LAK):
    - https://solaresearch.org/events/lak/
- Participate in a Learning Analytics Summer Institute (LASI):
    - https://solaresearch.org/events/lasi/
- Read and submit to our journal (JLA):
    - https://solaresearch.org/stay-informed/journal/

I hope you will enjoy and be stimulated by this volume and the others to come.

March, 2017

Stephanie D. Teasley, PhD

President, Society for Learning Analytics Research

Research Professor, University of Michigan

From the President of the International Educational Data Mining Society

Data science impacts many aspects of our life. It has been transforming industries, healthcare, as well as other sciences. Education is not an exception. On the contrary, an explosion of available data has revolutionized how much education research is done. An emerging area of educational data science is bringing together researchers and practitioners from many fields with an aim to better understand and improve learning processes through data-driven insights. Learning analytics is one of the new young disciplines in the area. It studies how to employ data mining, machine learning, natural language processing, visualization, and human-computer interaction approaches among others to provide educators and learners with insights that might improve learning processes and teaching practice.

A typical learning analytics question in the recent past would be how to predict accurately and early enough in the learning process whether a student in a course or in a study program is unlikely to complete it successfully. Since then, the landscape of learning analytics has been getting wider and wider as it becomes possible to track the behavior of learners and teachers in learning management systems, MOOCs, and other digital platforms that support educational processes. Of course, being able to collect larger volumes and varieties of educational data is only one of the necessary ingredients. It is essential to adopt, adapt and develop new computational techniques for analyzing it and for capitalizing on it.

The variety of questions that are being asked of the data are getting richer and richer. Many questions can be answered with well-established data mining techniques and other computational approaches, including but not limited to classification, clustering and pattern mining. More specialized educational data mining and learning analytics approaches, for instance Bayesian Knowledge Tracing, modeling student engagement, natural language analysis of learning discourse, student writing analytics, social network analysis of student interactions and performance monitoring with interactive dashboards are being used to get deeper insights into different learning processes.

I had the privilege of being among the first readers of the Handbook of Learning Analytics, before it was published. In my view the Handbook is a good first step for anyone wishing to get an introduction to learning analytics and to develop her conceptual understanding. The chapters are written in an accessible form and cover many of the essential topics in the field. The handbook would also benefit young and experienced learning analytics researchers, as well as ambassadors and enthusiasts of learning analytics.

The handbook goes beyond presenting learning analytics concepts, state-of-the-art techniques for analyzing different kinds of educational data, and selected applications and success stories from the field. The editors invited prominent researchers and practitioners of learning analytics and educational data mining to discuss important questions that are often not well understood by non-experts. Thus, some of the chapters invite readers to think about the inherently multidisciplinary research of learning analytics and about the value of at least considering, if not building on, established theories of learning sciences, educational psychology, and education research rather than going down the purely data-driven path. Other chapters emphasize the importance of understanding the difference between explanatory and predictive modeling or predictive and prescriptive analytics. This is especially useful for those who are responsible for, or simply thinking of, deploying learning analytics in an institution, or linking it with policy making processes or developing personalized learning ideas.

The future of learning analytics is bright. Educational institutions already see the many promising opportunities that it provides for advancing our understanding of different learning processes and enhancing them in a variety of educational settings. These include analytics-driven interventions in remedial education, advising students, better curriculum modeling and degree planning, and understanding student long-term success factors. We can witness how the potential of learning analytics is recognized not only by institutions, but also by companies developing educational software such as intelligent tutors, educational games, learning management systems or MOOC platforms. We can see that more and more of these tools include at least some elements of learning analytics.

Nevertheless, I would encourage the readers to think about the broad spectrum of ethical issues and accountability of learning analytics. We can collect valuable educational data and already have powerful tools for gaining insights into learning and the effectiveness of educational processes. We expand the set of questions asked to the data. Educational researcher and data scientists often have a false belief that computational approaches and off-the-shelf tools have no bad intent and if the tools are used correctly, then employing learning analytics is bound to produce success. It is critically important to educate early adopters of learning analytics not only about techniques and the kinds of findings from the data they facilitate, but also about their limits. The classical examples would remain the difference between correlation, predictive correlation and causation, and the interpretation of the statistical significance of the results. The Handbook provides a valuable educational support in this context too. Furthermore, authors of selected chapters share their thoughts and stimulate the discussion of several topics that have not been exhaustively studied in the field yet. I hope the Handbook will also help newcomers to realize the extent to which learning analytics can and should support educators and learners. Even more so, it is important to study in what ways learning analytics may mislead and potentially harm students, teachers and the ecosystems they are in and how to prevent this.

I would like to conclude with a forecast that learning analytics and educational data science will gain a deeper and more holistic understanding of what it means for learning analytics to be ethics-aware, accountable and transparent and how to achieve that. This will be instrumental for unlocking the full potential of learning analytics and driving the further healthy development of the field.

March 2017

Professor Mykola Pechenizkiy,

President of IEDMS – International Educational Data Mining Society

Chair Data Mining, Eindhoven University of Technology, the Netherlands

Table of Contents

SECTION 3. APPLICATIONS

SECTION 4. INSTITUTIONAL STRATEGIES & SYSTEMS PERSPECTIVES

SECTION 1

FOUNDATIONAL CONCEPTS

Chapter 1: Theory and Learning Analytics

Simon Knight, Simon Buckingham Shum

Connected Intelligence Centre, University of Technology Sydney, Australia
DOI: 10.18608/hla17.001

ABSTRACT

The challenge of understanding how theory and analytics relate is to move "from clicks to constructs" in a principled way. Learning analytics are a specific incarnation of the bigger shift to an algorithmically pervaded society, and their wider impact on education needs careful consideration. In this chapter, we argue that by design – or else by accident – the use of a learning analytics tool is always aligned with assessment regimes, which are in turn grounded in epistemological assumptions and pedagogical practices. Fundamentally then, we argue that deploying a given learning analytics tool expresses a commitment to a particular educational worldview, designed to nurture particular kinds of learners. We outline some key provocations in the development of learning analytic techniques, key questions to draw out the purpose and assumptions built into learning analytics. We suggest that using "claims analysis" – analysis of the implicit or explicit stances taken in the design and deploying of technologies – is a productive human-centred method to address these key questions, and we offer some examples of the method applied to those provocations.

Keywords: Theory, assessment regime, claims analysis

In what has become a well-cited, popular article in *Wired* magazine, in the new era of petabyte-scale data and analytics, Anderson (2008) envisaged the death of theory, models, and the scientific method. No longer do we need to create theories about how the world works, because the data will tell us directly as we discern, in almost real time, the impacts of probes and changes we make.

This high profile article and somewhat extreme conclusion, along with others (see, for example, Mayer-Schönberger & Cukier, 2013), has, not surprisingly, attracted criticism (boyd & Crawford, 2011; Pietsch, 2014).

Educational *researchers* are one community interested in the application of "big data" approaches in the form of learning analytics. A critical question turns on exactly how theory could, or should shape research in this new paradigm. Equally, a critical view is needed on how the new tools of the trade enhance/constrain theorizing by virtue of what they draw attention to, and what they ignore or downplay. Returning to our opening provocation from Anderson, the opposite conclusion is drawn by Wise and Shaffer (2015, p. 6):

> What counts as a meaningful finding when the number of data points is so large that something will always be significant? [...] In sum, when working with big data, theory is actually more important, not less, in interpreting results and identifying meaningful, actionable results. For this reason we have offered Data Geology (Shaffer, 2011; Arastoopour et al., 2014) and Data Archeology (Wise, 2014) as more appropriate metaphors than Data Mining for thinking about how we sift through the new masses of data while attending to underlying conceptual relationships and the situational context.

Data-intensive methods are having, and will continue to have, a transformative impact on scientific inquiry (Hey, Tansley, & Tolle, 2009), with familiar "big science" examples including genetics, astronomy, and high energy physics. The *BRCA2 gene*, *Red Dwarf* stars, and the *Higgs bosun* do not hold strong views on being computationally modelled, or who does what with the results. However, when *people* become aware that their behaviour is under surveillance, with potentially important consequences, they may choose to adapt or distort their behaviour to camouflage activity, or to game the system. Learning analytics researchers aiming to study learning using such tools must do so aware that they have adopted a particular set of lenses on "learning" that amplify and distort in particular ways, and that may unintentionally change the system being tracked. Researchers should stay alert to the emerging critical discourse around big data in society, data-intensive science broadly, as well as within education where the debate is at a nascent stage.

Let us turn now to *educators and learners*. The potential of learning analytics is arguably far more significant than as an enabler of data-intensive educational research, exciting as this is. The new possibility is that educators and learners – the stakeholders who constitute the learning system studied for so long by researchers – are for the first time able to see their own processes and progress rendered in ways that until now were the preserve of researchers outside the system. Data gathering, analysis, interpretation, and even intervention (in the case of adaptive software) is no longer the preserve of the researcher, but shifts to embedded sociotechnical educational infrastructure. So, for *educators and learners*, the interest turns on the ability to gain insight in a timely manner that could improve outcomes.

Thus, with people in the analytic loop, the system becomes reflexive (people change in response to the act of observation, and explicit feedback loops), and we confront new ethical dilemmas (Pardo & Siemens, 2014; Prinsloo & Slade, 2015). The design challenge moves from that of modelling closed, deterministic systems, into the space of "wicked problems" (Rittel, 1984) and complex adaptive systems (Deakin Crick, 2016; Macfadyen, Dawson, Pardo, & Gašević, 2014). As we hope to clarify, for someone trying to get a robust measure of "learning" from data traces, such reflexivity will be either a curse or a blessing, depending on how important learner agency and creativity are deemed to be, how fixed the intended learning outcomes are, whether analytical feedback loops are designed as interventions to shape learner cognition/interaction, and so forth.

Our view is that it is indeed likely that education, as both a research field and as a professional practice, is on the threshold of a data-intensive revolution analogous to that experienced by other fields. As the site of political and commercial interests, education is driven by policy imperatives for "impact evidence," and software products shipping with analytics dashboards. While such drivers are typically viewed with suspicion by educational practitioners and researchers, the opportunity is to be welcomed if we can learn how to harness and drive the new horsepower offered by analytics engines, in order to accelerate innovation and improve evidence-based decision-making. Systemic educational shifts are of course tough to effect, but could it be that analytics tools offer new ways to evidence, at scale, the kinds of process-intensive learning that educators have long argued for, but have to date proven impractical? Exactly what one seeks to do with analytics is at the heart of this chapter.

To design analytics-based lenses – with our eyes wide open to the risks of distorting our definition of "learning" in our desire to track it computationally – we must unpick what is at stake when classification schemes, machine learning, recommendation algorithms, and visualizations mediate the relationships between educators, learners, policymakers, and researchers. The challenge of understanding how theory and analytics relate is to move "from clicks to constructs" in a principled way.

Learning analytics are a specific incarnation of the bigger shift to an algorithmically pervaded society. The frame we place around the relationship of theory to learning analytics must therefore be enlarged beyond considerations of what is normally considered "educational theory," to engage with the critical discourse around how sociotechnical infrastructures deliver computational intelligence in society.

The remainder of the chapter argues that by design – or else by accident – the use of a learning analytics tool is always aligned with *assessment regimes*, which are in turn grounded in *epistemological assumptions* and *pedagogical practices*. Moreover, as we shall explain, a long history of design thinking demonstrates that designed artifacts unavoidably embody implicit values and claims. Fundamentally then, we argue that deploying a given learning analytics tool expresses a commitment to a particular educational worldview, designed to nurture particular kinds of learners.

THEORY INTO PRACTICE

In an earlier paper (Knight, Buckingham Shum, & Littleton, 2014) we put forward a triadic depiction of the relationship between elements of theory and practice in the development of learning analytic techniques, as depicted in Figure 1.1 (we refer the reader to this paper for further discussion of the depicted relationships). Our intention was to illustrate the tensions and inter-relations among the more or less theoretically grounded stances we take through our pedagogic and assessment practices and policies, and their underlying epistemological implications and assumptions.

The use of a triangle highlights these tensions: that assessment *can* be the driving force in how we understand what "knowledge" is; that assumptions about pedagogy (for example, a kind of folk psychology; Olson & Bruner, 1996) influence who we assess and how; that assessment and pedagogy are sometimes in tension, where the desire for summative assessment overrides pedagogically motivated formative feedback; and that drawing alignment between one's epistemological view (of the nature of knowledge) and assessment or pedagogy practices is challenging – relationships between the three may be implied, but they are not *entailed* (Davis & Williams, 2002). Of course, other visualizations might be imagined, and the theoretical and practical purposes for which such heuristics are devised is important to consider. To give two examples, we have considered versions of the depiction in which: 1) assessment and pedagogy are built on the foundation of epistemology (in a hierarchical structure), and 2) are brought into alignment in a Venn diagram structure, with greater overlap implying a greater complementarity of the theorized position.

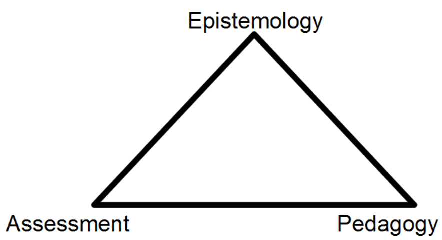


Figure 1.1. The Epistemology–Assessment–Pedagogy (EPA) triad (Knight et al., 2014, p. 25).

Learning analytics, as a new form of assessment instrument, have potential to support current educational practices, or to challenge them and reshape education; considering their theoretic-positioning is important in understanding the kind of education systems we aim for. For example, learning analytics could have the potential to 1) marginalize learners (and educators) through the transformation of education into a technocratic system; 2) limit what we talk about as "learning" to what we can create analytics for; and 3) exclude alternative ways of engaging in activities (that may be hard to track computationally), to the detriment of learners. Algorithms may both ignore, and mask some key elements of the learning process. The extent to which analytics can usefully support educators and learners is an important question. These are pressing issues given the rise of learning analytics, and increasing interest in mass online education at both the pre-university and university levels (e.g., the growing interest in MOOCs).

EPA PROVOCATIONS

Expanding on this prior work, the rest of the chapter aims to illustrate the application of our approach, with the aim of providing actionable guidance for those developing learning analytics approaches and tools. To do this, we have developed a set of provocations centred on the triad of epistemology, pedagogy, and assessment.

We use these provocations to illustrate how the implicit "claims" made by a learning analytics tool can be deconstructed. The approach invites reflection on the affordances of the tool's design at different levels (including data model, learner experience, and learning analytics visualization).

Computer-supported learning – individual or collaborative – covers a huge array of learning contexts. Such tools support many forms of rich learner interaction with peers and resources, which are implicit claims about learning. However, the emergence of computational analytics enables designers – and by extension the artifacts – to value certain behaviours above others, namely, those logged, analyzed, and rendered visible to some stakeholder group. The implicit claim is that these are particularly important behaviours. We measure what we value.

We provide a set of "six W" questions to be considered in the development of learning analytics. Of course, across these questions, there is overlap, and any one question might be expressed in multiple ways. The intention is neither to prescribe these as the only questions to be asked, nor that within each element of the triad *only* particular questions should apply. As the descriptions of the provocations make clear, within each facet of the triad, multiple theoretical questions can and should be asked. Rather, we hope to provide heuristic guidance to readers in developing their own analytics.

Epistemology – What Are We Measuring?

The first provocation invites the analytic designer to consider what "knowledge" looks like within the analytic approach being developed, asking, *What are we trying to measure*? We pose this question to prompt consideration of the connection between a conceptual account of the object of measurement (the knowledge being assessed), and a practical account of the methods and measures used to quantify activity and outputs within particular tasks. Asking *What are we trying to measure*? encourages us to consider our learning design, the skills and facts we want our students to learn, and what it means for students to "come to know." This is a question of epistemology; it concerns the nature of the constructs, why they "count" as knowledge, the evidentiary standard and kind required for a claim of knowledge to be made.

This knowledge might be of a more broadly propositional kind (sometimes characterized as "knowledge that," characterized as recall of facts), a more broad set of skills and characteristics (sometimes characterized as "knowledge how," for example, the ability to write an essay), or dispositions to act in particular ways (for example, as those dispositions recently discussed as epistemic virtues in epistemology). Evidentiary standards and types concern the warrants indicative of knowledge, for example, whether knowledge can be conceptualized in terms of unitary propositions that may be recalled more or less appropriately within particular contexts, whether knowledge of something entails the ability to deploy it in some context, the kinds of justification and warrant (and the skills underlying these) that cement some claim as knowledge, and so on. These are – implicitly or explicitly – the targets of our measurement.

Epistemology – How Are We Measuring?

Closely related to this conceptual question regarding the epistemological status of the object of analysis is a question regarding our access – as researchers and educators – to that knowledge. This is a question regarding the epistemological underpinning of our research and assessment methods. There is a rich literature on the various epistemological concerns around quantitative and qualitative research methods, with a growing specific interest in digital research methods. In addition, there is a focused literature in the philosophy of assessment, exploring the epistemological concerns in assessment methods (Davis, 1999). Across this literature, issues concerning the subjectivity of approaches, and the ability of methodologies to give insights, are central. The question invites considerations regarding the ways in which analytic methods imply particular epistemologies. Note that this is not just a question of the reliability of our assessment methods, but concerns the ability of approaches to speak to an externally knowable world (and the nature of that world).

Pedagogy – Why is this Knowledge Important to Us?

The development of analytic approaches in learning contexts involves making decisions about what knowledge will, and will not, be focused on; to measure what we value rather than value merely that which is easily measured (Wells & Claxton, 2002). This is, of course, in addition to a conceptual account of the nature of that knowledge. These decisions in part relate to debates around the kinds of important (or powerful) knowledge in society (see, for example, Young & Muller, 2015) and the role of knowledge-based curricula, including discussions around employability (or the balance of vocational and liberal educational aims), 21st-century skills, and so on. This question asks, *Why does this analytic matter to educators and learners*?

Answering this question might in part be salient to the kind of learning theory that the analytic sits within; to instrumental aims regarding the analytic's contribution to particular skills (perhaps employability skills); or university compliance (for example, reporting requirements). It might also relate to pedagogic aims such as the support of particular groups of students, and so on.

Pedagogy – Who is the Assessment/ Analytic For?

Extending the concern with the nature of the object of assessment above, is a further concern regarding the target of the analytic device, provoking the question, *Who is the analytic for*? In the development of analytic devices (and assessments more broadly), we should consider who the target of the device is, whether it supports teachers, parents, students, or administrators in understanding some aspect of learning. Is the analytic designed to provide insight at a macro (government, institutional), meso (school, class), or micro (individual student or activity) level (Buckingham Shum, 2012), and are there insights across these levels that can be effectively made sense of by all stakeholders (Knight, Buckingham Shum, & Littleton, 2013)?

This question regards the desire for analytic insights at multiple levels of a system, and the ability of individual analytic approaches (including their outputs in various forms, such as dashboards) to support the following: 1) individual students in developing their learning; 2) educators in developing their own practice and in targeting their support at individual student needs; and 3) administrators in understanding how cohorts are developing and their organizational needs. As Crick argues in this handbook, a complex systems conception of analytics for different levels in the learning system, spanning from private personal data through to shared organizational data, implies different rationalities and authorities to interpret at the different levels (Deakin Crick, 2017).

This question raises a parallel concern regarding the ethical implications in developing analytics that (explicitly or implicitly) target particular groups. This concern is at two levels. First, analytics that require particular forms of technology or participation may create new divides between student cohorts, or entrench existing divides. Second, there is an ethical concern regarding the use of student data by institutions, particularly where specific consent is not given, where no direct learning gain is directed to those students. This second issue is a particular consideration in cases wherein student data may be used largely to reduce institutional costs or the level of support given to particular students.

Assessment – Where Does the Assessment Happen?

Obvious though this is, we note that assessment always takes place in a physical location, in response to particular task demands, in a sociocultural context, with a particular set of tools. Contrast an individual pen-and-paper exam in a silent hall with 300 peers, with an emergency response simulation on the ward, with tackling a statistics problem in a MOOC, with conflict resolution in relationship counselling. For each context, we must ask not only if the assessment is meaningful, but also to what extent meaningful computational analytics can be designed to add value.

Moreover, we should also consider the ways in which the assessment biases particular kinds of response – in sometimes-unintended ways. For example, particular groups of students, or kinds of knowledge, might be privileged over others through the design of assessment contexts with a very narrow definition of achievement; through requirements for behaviours that not all students might engage in; through the use of technologies that unfairly assume socioeconomic means; or through separating assessment from the applied context in which an expertise can be displayed authentically.

Across assessments, we should also consider the ways in which the particular systems shape the data obtained – note this is a practical concern regarding the reliability and validity of methods, rather than the related epistemological concern raised above. For example, technologies are mediational tools, which shape the ways in which people interact with each other and the world around them, and hence, the activity they measure. This is true both of the specific technologies, and the task design used in assessments; for example, the use of "authentic" assessments provides a different range of possible responses than more traditional pen-and-paper assessments of various kinds.

Assessment – When Does the Assessment, and Feedback, Occur?

A final consideration relates to the temporal context of learning analytics, asking when the assessment and feedback cycle occurs. This provocation is intended to prompt consideration of the formative or/and summative nature of the learning analytic; whether or not a particular technology provides after-the-fact or real-time feedback, and whether this feedback is intended to provide a scaffold or model for current behaviour, is targeted at future behaviour and learning, or is just intended as a feedback mechanism on prior work (which may not be covered again).

In our earlier paper, we made a distinction between the metaphors of biofeedback and diagnostic learning analytics. The intention here was to draw a distinction between formative and summative assessments (respectively). However, while the analogy can be drawn, of course systems that provide real-time feedback can be summative in nature, and can take on a "monitoring" role towards some end-point. In addition, diagnosis does not have the finality perhaps implied by the analogy – diagnosis provides insight into what is going wrong, which may be actionable by the "patient" (student) or "doctor" (educator). Instead, then, the focus should be on whether the analytic device is targeted at a summative snapshot perspective on student learning and monitoring towards that end, or instead, targeted at development and improvement over time.

CONCLUSION

Through the provocations, we have drawn attention to the ways in which analytic approaches and artifacts subscribe to particular perspectives on learning: they implicitly *make claims* across the EPA triad, and the provocations drawn from them.

Tools can be used in many ways, and should not be isolated from their context of use. Interactional affordances, like beauty, are to some degree "in the eye of the beholder." We offer these provocations as a pragmatic tool for thinking, for designers, educators, researchers, and students – whether considering how one currently makes use of analytical tools, how one might do in the future, or indeed when designing new tools for new contexts. We propose that it is productive to consider these provocations in order to reflect on the EPA claims being made through the deployment of a learning analytic tool within a given context.

REFERENCES

Anderson, C. (2008, 23 June). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Arastoopour, G., Buckingham Shum, S., Collier, W., Kirschner, P., Knight, S. K., Shaffer, D. W., & Wise, A. F. (2014). Analytics for learning and becoming in practice. In J. L. Polman et al. (Eds.), *Learning and Becoming in Practice: Proceedings of the International Conference of the Learning Sciences* (ICLS '14), 23–27 June 2014, Boulder, CO, USA (Vol. 3, pp. 1680–1683). Boulder, CO: International Society of the Learning Sciences

boyd, danah, & Crawford, K. (2011, September). *Six provocations for big data*. Presented at *A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*. http://ssrn.com/abstract=1926431

Buckingham Shum, S. (2012). *Learning analytics policy brief*. UNESCO. http://iite.unesco.org/pics/publications/en/files/3214711.pdf

Davis, A. (1999). *The limits of educational assessment*. Wiley.

Davis, A., & Williams, K. (2002). Epistemology and curriculum. In N. Blake, P. Smeyers, & R. Smith (Eds.), *Blackwell guide to the philosophy of education*. Blackwell Reference Online.

Deakin Crick, R. (2017). Learning analytics: Layers, loops, and processes in a virtual learning infrastructure. In C. Lang & G. Siemens (Eds.), *Handbook on Learning Analytics* (this volume).

Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research. http://research.microsoft.com/en-us/collaboration/fourthparadigm/

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Collaborative sensemaking in learning analytics. CSCW *and Education Workshop (2013): Viewing education as a site of work practice, co-located with the 16th ACM Conference on Computer Support Cooperative Work and Social Computing* (CSCW 2013), 23 February 2013, San Antonio, Texas, USA. http://oro.open.ac.uk/36582/

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space. *Journal of Learning Analytics*, 1(2). http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/3538

Macfadyen, L. P., Dawson, S., Pardo, A., & Gašević, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, 9, 17–28.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live work and think*. London: John Murray Publisher.

Olson, D. R., & Bruner, J. S. (1996). Folk psychology and folk pedagogy. In D. R. Olson & N. Torrance (Eds.), *The Handbook of Education and Human Development* (pp. 9–27). Hoboken, NJ: Wiley-Blackwell.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. http://doi.org/10.1111/bjet.12152

Pietsch, W. (2014). Aspects of theory-ladenness in data-intensive science. Paper presented at the 24th Biennial Meeting of the Philosophy of Science Association (PSA 2014), 6–9 November 2014, Chicago, IL. http://philsci-archive.pitt.edu/10777/

Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March, Poughkeepsie, NY, USA (pp. 83–92). New York: ACM.

Rittel, H. (1984). Second generation design methods. In N. Cross (Ed.), *Developments in design methodology* (pp. 317–327). Chichester, UK: John Wiley & Sons.

Shaffer, D. W. (2011, December). Epistemic network assessment. Presentation at the National Academy of Education Summit, Washington, DC.

Wells, G., & Claxton, G. (2002). *Learning for life in the 21st century: Sociocultural perspectives on the future of education*. Hoboken, NJ: Wiley-Blackwell.

Wise, A. F. (2014). Data archeology: A theory informed approach to analyzing data traces of social interaction in large scale learning environments. *Proceedings of the Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses at the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP 2014), 25 October 2014, Doha, Qatar (pp. 1–2). Association for Computational Linguistics. https://www.aclweb.org/anthology/W/W14/W14-4100.pdf

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, 2(2), 5–13.

Young, M., & Muller, J. (2015). *Curriculum and the specialization of knowledge: Studies in the sociology of education*. London: Routledge.

Chapter 2: Computational Methods for the Analysis of Learning and Knowledge Building Communities

H. Ulrich Hoppe

Department of Computer Science and Applied Cognitive Science, University of Duisburg-Essen, Germany
DOI: 10.18608/hla17.002

ABSTRACT

Learning analytics (LA) features an inherent interest in algorithms and computational methods of analysis. This makes LA an interesting field of study for computer scientists and mathematically inspired researchers. A differentiated view of the different types of approaches is relevant not only for "technologists" but also for the design and interpretation of analytics applications. The "trinity of methods" includes analytics of 1) *network structures* including actor–actor (social) networks but also actor–artefact networks, 2) *processes* using methods of sequence analysis, and 3) *content* using text mining or other techniques of artefact analysis. A summary picture of these approaches and their roots is given. Two recent studies are presented to exemplify challenges and potential benefits of using advanced computational methods that combine different methodological approaches.

Keywords: Trinity of computational methods, knowledge building, actor–artefact networks, resource access patterns

The newly established field of learning analytics (LA) features an inherent interest in computational or algorithmic methods of data analysis. In this perspective, "analytics" is more than just the empirical analysis of learning interactions in technology-rich settings, it actually also calls for specific computational and mathematical approaches as part of the analysis. This line of research builds on techniques of data mining and network analysis, which are adapted, specialized, and potentially developed further in an LA context.

To better understand the potential and challenges of this endeavor, it is important to introduce some distinctions regarding the nature of the underlying methods. Computational approaches used in LA include analytics of 1) *network structures* including actor–actor (social) networks but also actor–artefact networks, 2) *processes* using methods of sequence analysis, and 3) *content* using text mining or other techniques of computational artefact analysis. This distinction is not only relevant for "technologists" who actually work with and on the computational methods, it is also important for the design of "LA-enriched" educational environments and scenarios. We should not expect LA to develop new computational-analytic techniques from scratch but to adapt and possibly extend existing approaches in an LA context. First, the different premises and affordances of the different types of methods should be well understood. Furthermore, the combination and synergetic use of different types of methods is often desirable from an application perspective but this constitutes new challenges from a conceptual as well as a computational point of view.

The computational analysis of interaction and communication in group learning scenarios and learning communities has been a topic of research even before the field of LA was constituted, and this work is still relevant to LA. Early adoptions of social network analysis (SNA type 1) in this context include the works of Haythornthwaite (2001), Reffay & Chanier (2003), Harrer, Malzahn, Zeini, and Hoppe (2007), and De Laat, Lally, Lipponen, and Simons (2007). Process-oriented analytics techniques (type 2) have an even longer history, especially in the analysis of interactions in a computer-supported collaborative learning (CSCL) context (Mühlenbrock & Hoppe, 1999; Harrer, Martínez-Monés, & Dimitracopoulou, 2009). Although somewhat later and less numerous, content-based analyses (type 3) based on computational linguistics techniques have been successfully applied to the analysis of collaborative learning processes (e.g., by Rosé et al., 2008).

Network-Analytic Methods

Network-analytic approaches, especially social network analysis (SNA), are characterized by taking a relational perspective and by viewing actors as nodes in a network, represented as a graph structure. In this sense, a network consists of a set of actors, and a set of ties between pairs of actors (Wasserman & Faust, 1994). The type of pairwise connection defines the nature of each social network (Borgatti, Mehra, Brass, & Labianca, 2009). Examples of different types of ties are affiliation, friendship, professional, behavioural interaction, or information sharing. The visualization of such network structures has emerged as a specific subfield (Krempel, 2005). Standard methods of network analysis allow for quantifying the importance of actors by different types of "centrality measures" and detecting clusters of actors connected more densely among each other than the average (detection of "cohesive subgroups" or "community detection" – for an overview, see Fortunato, 2010).

A well-known inherent limitation of SNA is that the target representation, i.e., the social network, aggregates data over a given time window but no longer represents the underlying temporal dynamics (i.e., interaction patterns). It has been shown that the size of the time window of aggregation has a systematic influence on certain network characteristics such as subcommunity structures (Zeini, Göhnert, Hecking, Krempel, & Hoppe, 2014). To explicitly address time-dependent effects, SNA techniques have been extended to analyzing time series of networks in dynamic approaches.

It is important to acknowledge that network analytic techniques (even under the heading of SNA) do not exclusively deal with actors and social relations as basic elements. So-called "affiliation networks" or "two-mode networks" (Wasserman & Faust, 1994) are based on relations between two distinct types of entities, namely actors and affiliations. Here, the "affiliation" type can be of a very different nature, including, for example, publications as affiliations in relation to authors as actors in the context of coauthoring networks. In general, two-mode networks can be used to model the creation and sharing of knowledge artefacts in knowledge building scenarios. In pure form, these networks are assumed to be bipartite, i.e., only alternating links actor–artefact (relation "created/modified") or artefact–actor (relation "created-by/modified-by") would be allowed. Using simple matrix operations, such bipartite two-mode-networks can be "folded" into homogeneous (one-mode) networks of either only actors or only artefacts. Here, for example, two actors would be associated if they have acted upon the same artefact. We would then say that the relation between the actors was mediated by the artefact. Similarly, we can derive relationships between artefacts by considering agents (one actor engaged in the creation of two different artefacts) as mediators.

We have seen an increasing number of studies of educational communities using SNA techniques related to networked learning and CSCL. Originally, networks derived from email and discussion boards were the most prominent conditions studied, as for example the early study of cohesion in learning groups (Reffay & Chanier, 2003). Meanwhile, network analysis belongs to the core of LA techniques. The classification of approaches to "social learning analytics" by Ferguson and Shum (2012), though not primarily computationally oriented, prominently mentions network analysis techniques including both actor–actor and actor–artefact networks.

Process-Oriented Interaction Analysis

The computational analysis of learner (inter-)actions based on the system's logfiles has a tradition in CSCL. There were even attempts to standardize action-logging formats in CSCL systems to facilitate the sharing and combination of existing interaction analysis techniques (Harrer et al., 2009). One of the earliest examples of applying intelligent computational techniques in a CSCL context (namely sequential pattern recognition) was suggested and exemplified by Mühlenbrock and Hoppe (1999). This approach was later used in an empirical context to pre-process CSCL action logs in order to automatically detect the occurrence of certain collaboration patterns such as "co-construction" or "conflict" (Zumbach, Mühlenbrock, Jansen, Reimann, & Hoppe, 2002).

Whereas these approaches were developed in a learning-related research context, there are also more general techniques that can be adapted and used, such as the scalable platforms management forum (SPMF) and library of sequential patterns mining methods (Fournier-Viger et al., 2014). In an LA context, SPMF is used by the LeMo tool suite for the analytics of activities on online learning platforms (Elkina, Fortenbacher, Merceron, 2013). In another recent study, Bannert, Reimann, & Sonnenberg (2014) have used "process mining," a computational technique with roots in automata theory, to characterize patterns and strategies in self-regulated learning.

Content Analysis Using Text-Mining Methods

There is a tradition of content-analysis-based human interpretation and coding often used as input to quantitative empirical research, as discussed by Strijbos, Martens, Prins, and Jochems (2006) from a CSCL perspective. In contrast, from a computational point of view, we are interested in applying information-mining techniques to extract semantic information from artefacts. Obviously, this is of particular interest in the case of learner-generated artefacts. Rosé et al. (2008) has demonstrated the usefulness of automatic text classification with a corpus of CSCL transcripts. Sherin (2013) used computational techniques of content analysis on student interview data to discover the students' understanding of science concepts. Content analysis techniques have also been used for the clustering of e-learning resources according to their similarity (Hung, 2012). He (2013) proposed the usage of similar techniques for grouping learners' main topics in student-to-teacher online questions and peer-to-peer chat messages in the context of online video-based learning.

Typically, these methods of textual content analysis are based on the "bag of words" model in which the given order of words in a text is of no relevance to the analysis. This is the case for a variety of probabilistic topic modelling techniques such as the currently quite popular method of latent Dirichlet allocation (LDA; Blei, 2012). A method that does take into account the positioning of words in a text is network text analysis (NTA). NTA is a text mining method that connects content analysis with network representations in that it extracts a network of concepts from given texts (Carley, Columbus, & Landwehr, 2013). Links between concepts are established if the corresponding terms co-occur with a certain frequency in a sliding window of pre-specified width that runs over a normalized version of the text. A "meta thesaurus" allows for introducing different concept categories (e.g., "person," "location," "domain_concept" et cetera). On this basis, multimode networks can be formed, in which the concept–concept relations are restricted to certain inter-category types such as l*ocation–person* or *person–domain_concept*. These representations can in turn be analyzed using network-analytic concepts such as centrality measures or the detection of cohesive subgroups as a network-based clustering technique.

Figure 2.1 shows the result of applying NTA to transcripts from teacher–student workshops in the context of the European project JuxtaLearn (Hoppe, Erkens, Clough, Daems, & Adams, 2013). The resulting networks nicely reflect the different topics from the areas of biology, chemistry, and physics, initially presented by students and then discussed in the whole group. Here *topics* (pentagon-shaped nodes) and *topic–topic* relations are depicted in grey, whereas *persons* (square nodes) and *person–topic* relations are darker (black).

In the topic–topic network, the three different fields of science appear as more densely connected islands (or "cohesive subgroups") although certain cross-links exist (e.g., *diffusion* in biology is linked to *molecule* in chemistry). The person–topic links allow for judging the importance of the individuals' contributions in the presentations and discussion. Most contributions stay within one subfield. Student S5 stands out in terms of degree centrality (14 connections to different topics) and with most contributions to physics but one link bridging over into chemistry. In the JuxtaLearn project, this approach has been further developed to assess students' problems of understanding from question/answer collections related to science videos (Daems, Erkens, Malzahn, & Hoppe, 2014). The extracted net-

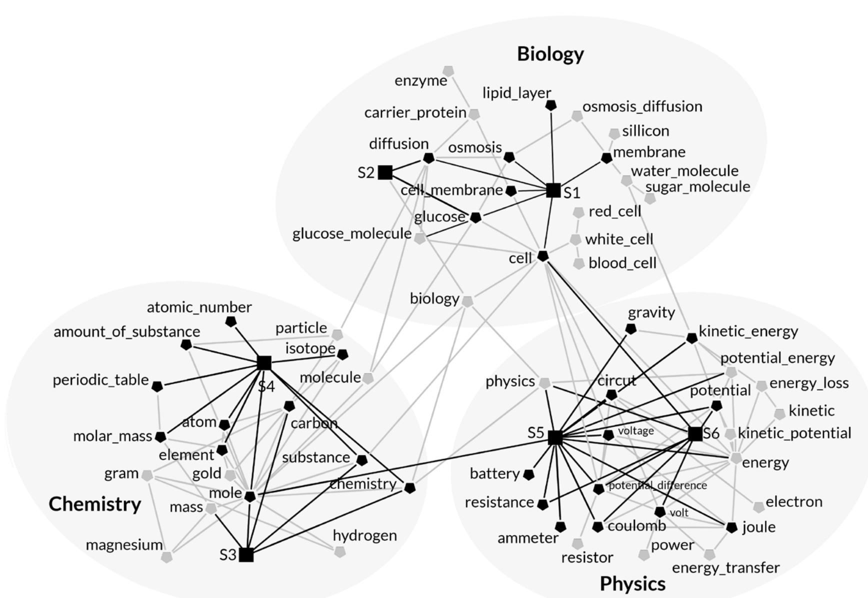


Figure 2.1. Topic–topic and person–topic relations extracted from transcripts of teacher–student workshops.

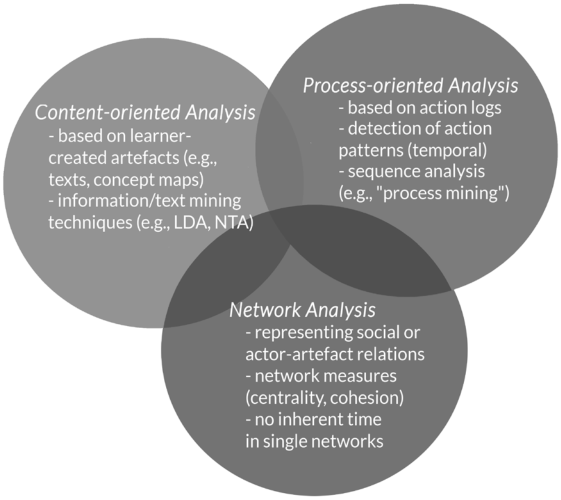


Figure 2.2. The "trinity" of methodological approaches.

works of concepts have been contrasted with teacher-created taxonomies. This has led to an enrichment of the taxonomies and to the identification of specific problems of understanding on the part of the learners. From a pedagogical perspective, this provides empirical insights relevant to curriculum construction and curriculum revision (here specifically related to teacher-created micro-curricula).

Figure 2.2 summarizes the characteristics of the three methodological approaches in terms of their basic representational characteristics and typical techniques. Overlapping areas between the approaches are of particular importance for new integrative or synergetic applications.

The remainder of this article presents two case studies of applying specific computational techniques to the analysis of learning and knowledge building in communities. The first example shows that more sophisticated methods of network analysis may yield interesting insights in a case where "first order approaches" would fail to resolve interesting structures. In that it considers the evolution of patterns of resource access on a learning platform over time, it combines the network analytic approach with process aspects. The second case describes the adoption and revision of a scientometric method to characterize the evolution of ideas in a knowledge-building community. This network-analytic approach is then combined with content-based text mining methods. So, both examples support the general point that we can expect additional benefit from combining different methods.

Example 1: Dynamic Resource Access Patterns in Online Courses

Nowadays higher education practice is commonly supported by learning platforms such as Moodle to distribute educational materials of various types, including lecture slides, videos, and task assignments, but also to collect exercise or quiz results and to facilitate individual or group work using forums or wikis. In this way, classical presence lectures are turned into blended learning courses or, according to Fox (2013), "small private online courses" (SPOCs). As for the traces that learners leave on such learning platforms, the most abundant actions are resource access activities that constitute actor (learner) – artefact (learning resource) relations. Only in special cases, such as the co-editing of wiki articles, such data may be interpreted in a quite straightforward way as actor–actor relations by "folding away" the mediating artifact (i.e., inter-connecting co-authors of the same wiki article). If applied to instructor-provided lecture materials, the actor–actor relation based on access to the same lecture would not be selective and would result in a dense network. Accordingly, the detection and tracing of clusters or subcommunities in such induced actor–actor networks would not be likely to provide interesting insights.

In a study based on one of the author's regular master courses (Hecking, Ziebarth, & Hoppe, 2014), a more sophisticated technique has been used to overcome this problem. Applying a subcommunity detection algorithm for two-mode networks to the original learner-resource data leads to much more selective and differentiated results in terms of identifying groups of learners working with certain groups of materials in a given time slice. This approach is based on the network-analytic method of "bi-clique percolation analysis" (Lehmann, Schwarz, & Hansen, 2008), which is a generalization of the clique percolation method originally defined for one-mode networks (Palla, Derenyi, Farkas, & Vicsek, 2005). The clique percolation method (CPM) builds subcommunities on the presence of cliques (fully connected subgraphs) in one-mode networks. CPM is of particular interest for the analysis of collaborating communities because the resulting clusters may overlap and thus can also be used to identify potential brokers or mediators between different subcommunities. This characteristic also holds for the bi-clique percolation method (BCPM) with two-mode networks. In their original article, Lehmann et al. (2008) identified the higher selectivity of subcommunity in the two-mode network. We have been able to corroborate this in our application case.

Figure 2.3 shows how, on principle, BCPM can be used to trace cohesive clusters in two-mode networks. First, BCPM is applied to each time slice of the network (left-hand side). The diagram on the right abstracts from the individual entities and just depicts and inter-links between groups of actors (squares) and groups of resources (circles). In one particular time slice, two

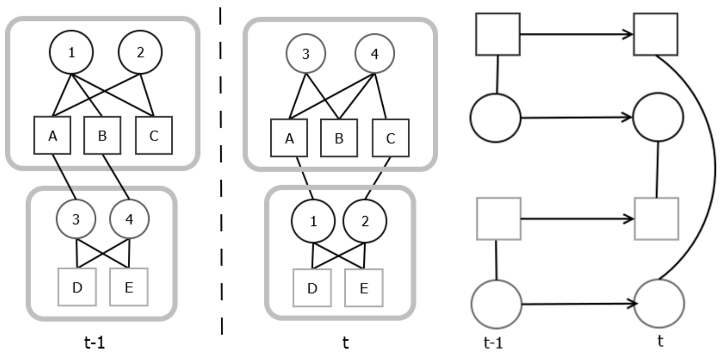


Figure 2.3. Evolving two-mode clusters (left) and the corresponding swim lane diagram (right).

groups of different types are linked by a vertical edge, indicating that these two groups form a bipartite cluster. Horizontal edges appear across time slices and link two groups of the same type, indicating that the two groups can be considered as similar. Here, we see a situation where the connection between actors and resources is switched from one time slice to other. In general, it is not clear if the basic groups "survive" from one time slice to the other (as is the case here). Palla, Barabasi, and Vicsek (2007) have defined a complex system of transformations (such as "birth," "merger," "split" et cetera) that can be used to trace the evolution of subgroups over time.

In our study (Hecking et al., 2014), affiliation networks were built based on students' access to learning resources during a blended learning course on interactive learning and teaching technologies. The course was resource intensive in the sense that the traditional lecture was accompanied by a variety of additional learning resources like lecture videos, slides, serious games, as well as a glossary of important concepts created by the students themselves as a wiki. Students and resources were simultaneously grouped into mixed and overlapping clusters as explained above. Those clusters can be interpreted as a group of students who have a common interest in a group of learning resources but not necessarily having social connections. A typical example cluster is depicted in Figure 2.3.

By applying the method to the student–resource networks of particular weeks during the lecture period, this analysis reflects certain groupings induced by explicit assignments but also yields some surprising insights regarding the usage materials. This can be seen, for example, in Figure 2.4 where the orange coloured cluster comprises lecture videos and students who seem to have a distinct interest in learning resources compared to the others.

In addition, the tracking of bipartite student–resource clusters was used to investigate student resource-access behaviour during exam preparation after the last lecture. This period is particular interesting because by then the entire pool of learning materials, successively added week by week to the course, was available, including the wiki articles created by the students.

The swim lane diagram in Figure 2.5 depicts the resource access patterns found in the course during this phase. Time slices where build based on a time window size of 4 days. The oral exams were distributed over two weeks for most of the students, while for another study program the examination phase began six weeks after the last lecture. One finding is that a large majority of students accessed large portions of the learning material over several time slices (highlighted blue box). Between time slices 2 and 5, there was a stable

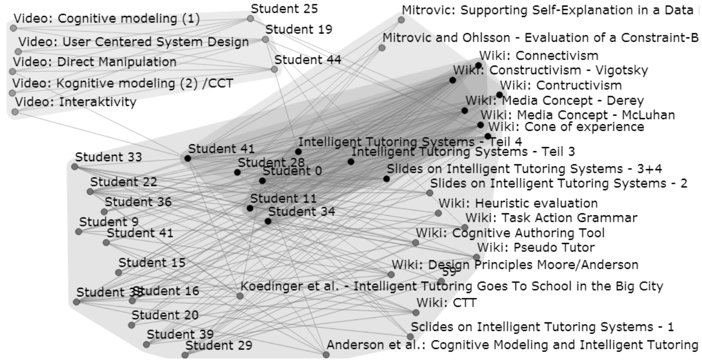


Figure 2.4. Bipartite clusters of students and learning resources (black nodes belong to more than one cluster).

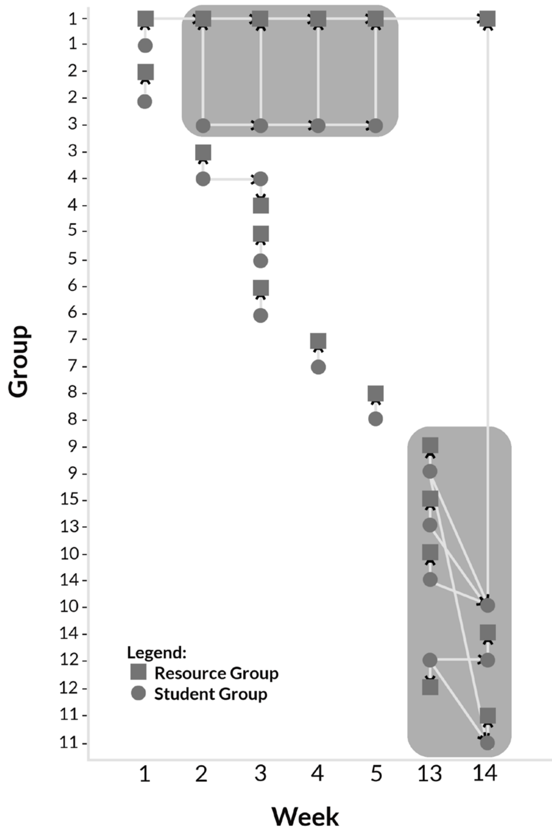


Figure 2.5. Swim lane diagram of the evolving student–resource clusters during the exam phase.

set of students (stud. group 3) using this material for exam preparation. In contrast, the students of the study program who had their oral exams later had a more diverse resource access behaviour (green box). Also, they began their exam preparation much closer to the time of the exam compared to the other study programs. In the last time slice, three of the four student groups merged to a larger group that was then more affiliated to the core learning material (res. group 1).

On the one hand, this example shows the possible expressiveness of sophisticated network analysis methods in a case where "first order methods" would not be able to resolve interesting and meaningful structural relations. On the other hand, it demonstrates that additional effort is needed to support a dynamic, evolutionary interpretation of network-based models (given that each single network is "ignorant" about time).

In an ongoing research project on supporting small group collaboration in MOOCs, we have used this approach of tracking cohesive clusters of learners and resources to distinguish "mainstreaming behaviour" from more individual or idiosyncratic patterns of resource usage on the part of learners (Ziebarth et al., 2015). Given this model-based distinction, we found that extrinsic motivation was more prevalent in the mainstreaming group. This suggests that specific patterns in actor–artefact relations may serve as indicators for learning styles.

Example 2: Analyzing the Evolution of Ideas in Knowledge-building Communities

Scientific production can be seen as a prototypical case of knowledge building in a community. Accordingly, methods developed to analyze scientific production and collaboration ("scientometric methods") can plausibly also be used to analyze other types of knowledge building in networked learning communities. Hummon and Doreian (1989) have proposed the method of "main path analysis" (MPA) to detect the main flow of ideas in citation networks with scientific publications as nodes connected by citations. The original paper uses a corpus of publications in DNA biology as an example.

The MPA method relies on the acyclic nature of citation graphs. Different from other network-analytic techniques, MPA has an implicit notion of time that stems from the nature of citation networks (always the citing paper is more recent than the cited one). As a consequence of this time ordering, and given that every collection is finite, in a corpus, there are always documents not cited by others (end-points or "sinks") as well as documents that do not cite other documents in the corpus ("sources"). The idea of MPA is to find the most used edges in terms of the information flow from the source nodes to the sink nodes. One common approach to finding these edges is the "search path count" or SPC method (Batagelj, 2003). All sources in the network are connected to a single artificial source and all sinks to a single artificial sink. SPC assigns a weight to an edge according to the number of paths from the source to the sink on which the edge occurs. The main path can then be found by traversing the graph from the source to the sink by using the edges with the highest weight, as depicted in Figure 2.6.

The idea of applying MPA to learning communities working with hyper-linked connections of wiki documents was first proposed by Iassen Halatchliyski and colleagues (2012). However, MPA cannot be applied directly to hyper-linked web documents because the premise of directed acyclic graphs (DAGs) is usually not fulfilled. Since the content of articles in a wiki is dynamically evolving, hyperlinks between two articles do not induce a temporal order between them and cycles or even bidirectional citation links are quite frequent. In Halatchliyski, Hecking, Göhnert, & Hoppe (2014), we have proposed a formal modification that allows for applying MPA also to this case. The adapted approach considers the particular revisions (successive versions) of articles instead of the articles themselves. Revisions of an evolving wiki article are artefacts with stable content as scientific publications. In such a network based on versions as nodes, we introduce revision edges between successive revisions of the same article. The original hyperlinks between different articles connect specific revisions and also introduce new versions. This trick avoids cyclic structures and allows for applying MPA. In the context of the Wikiversity learning community, we have used the coincidence of articles with identified main paths has as a basis to judge the importance or weight of contributions and to characterize author profiles in terms of specific role models (inspirator, connector, worker). These characterizations serve as supportive information for the management of knowledge building communities.

In a subsequent study (Hoppe, Göhnert, Steinert, & Charles, 2014), we have combined the network-analytic method of MPA with content analyses to analyze chat interactions in an educational community (*Tapped In* – see Farooq, Schank, Harris, Fusco, & Schlager, 2007). Here, the characteristic of chat as a synchronous communication medium, especially regarding turn taking, possible parallel threading, and interactional coherence had to be taken into account. Our work used contingency analysis (Suthers, Dwyer, Medina, & Vatraou, 2010) as theoretical background and reference to detect general dependencies based on operational

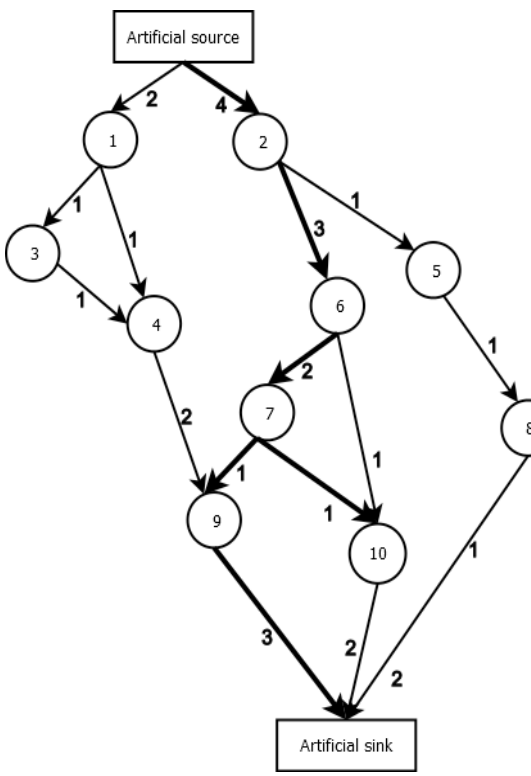


Figure 2.6. Example network illustrating the SPC method (edge weights are SPC counts; thick edges indicate the main paths).

rules. We reconstructed and refined this approach by using the concept of dialogue act tagging (Wu, Khan, Fisher, Shuler, & Pottenger, 2005) to enrich the basic set of indicators. We have tested our method using several examples of chat protocols from a teacher community as benchmarks. This allowed us to assess the agreement between the contingency links generated by our method with previously hand-coded contingencies (Suthers & Desiato, 2012) based on the F-score (a measure used in information retrieval combining precision and recall). The automatically generated contingencies reached an F-score similarity of 83% to 97%, which is comparable to the pairwise F-score similarity of manually analyzed graphs. Figure 2.7 shows a fragment of a chat sequence with contingency links indicated on the right hand side, main path contributions highlighted in bold, and the message categories resulting from dialogue act tagging (e.g., "Statement" or "ynQuestion") added in brackets.

The main path information should be interpreted as an indicator for the relevance of contributions in the evolution and progress of the overall discourse. This relevance measure for contributions can in turn be used to estimate the influence of participants in the discourse. Since we did not have human ratings for this feature, we have compared the measure "percentage of contributions on main path" (*%MainPath*) per actor to other influence rankings based on the well-established *PageRank* and *Indegree* measures. We applied these measures to different versions of the contingency graphs resulting from human and automatic coding. As a result, we found a 0.82 (0.82) correlation of *%MainPath* with *PageRank* and a 0.69 (0.88) correlation with *Indegree*. Per se, *%MainPath* is just another competing indicator. However it is different from the other measures since it takes into account the flow of arguments in the discourse and not only local (*Indegree*) or globally weighted prestige (*PageRank*). As can be seen in Figure 2.6, MPA also allows for filtering the discourse for main threads of argument. In this sense, MPA makes the network model more specific and meaningful. However, further investigation is needed to validate these constructs.

DISCUSSION AND OUTLOOK

In general, we cannot expect LA to invent genuinely new computational methods of data mining and analysis. Yet, we have seen the successful adoption of a number of existing techniques. A prominent case is certainly social network analysis – to the extent that SNA concepts such as centrality measures or cohesive clusters (subcommunities) are now part of the conceptual repertoire used in LA discourse. This is still less the case for process-oriented techniques (such as

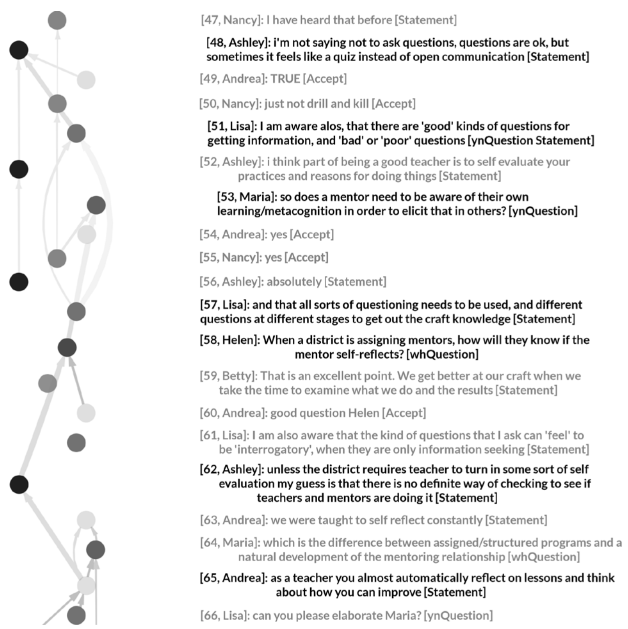


Figure 2.7. Fragment of a chat protocol with inferred contingency links (mainpath contributions and links indicated in bold).

sequential pattern mining) or linguistics-based methods for the analysis of dialogue and textual artefacts. The examples and arguments presented in this article corroborate 1) that even SNA has more to offer than the better known "first order approaches" and 2) the most benefit can be expected from combining different types of analytic methods. Regarding network analysis techniques, moving from pure actor–actor networks to actor–artefact (or two-mode) networks provides a richer basis of information that can resolve more significant and meaningful relations. The example on analyzing resource access in online courses illustrates how this can make a difference. It also shows the inclusion of time by considering temporal sequences of networks.

This article looks at the issues and challenges primarily from a computational perspective. From a pedagogical perspective, it is important to identify the affordances but also the deficits of certain methods in order to judge their potential benefits. For example, when targeting "interaction patterns" in knowledge building communities it should be clear that pure SNA models would only reveal static actor–actor relations but not time-dependent patterns. Possible extensions would use time series of networks and/or actor–artefact relations. Network-text analysis is an example of an approach that converts textual artefacts into networks of concepts (of possibly different categories) and thus allows for combining content and network analytic approaches. On the other hand, given these computational methods, where are the potential pedagogical added values? In this respect, we have seen the following examples:

- Concept networks derived from learner-generated texts using NTA can reveal students' mental models and misconceptions. This can be a basis for enriching domain taxonomies and for curriculum revision.
- The primary information that we get from learning platforms is about learners accessing (or possibly creating/uploading) resources. From sequences of ensuing two-mode learner–artefact networks, we can classify learner behaviour as "main-streaming" or more individually varied, possibly intrinsically motivated or curiosity-driven.
- Techniques borrowed from scientometrics allow for identifying the main lines of the evolution of ideas in knowledge building communities. On this basis, we can characterize contributions and the role of contributors to support better-informed decisions in the management of the community.

Forum participation in massive online courses has recently been the subject of several LA-inspired studies. Using a mix of analytic techniques involving SNA patterns combined with "regularity" of interactions and content assessment (based on human ratings), Poquet and Dawson (2016) have characterized success factors for productive and supportive forum interactions. Interestingly, they found an important influence of certain community members who were not themselves part of densely connected subgroups (or "cliques") on the positive evolution of the networked community. In a similar context, Wise, Cui, and Vytasek (2016) have identified certain linguistic features as predictors for distinguishing content-related from non-content-related (social or organizational) talk in such forums. In addition to domain-specific vocabulary, they also found general terms such as "understand," "example," "difference," or question words among the predictors of content-related contributions. These, in turn, correspond to so-called "signal concepts" used in the network-text analysis of educational video comments by Daems, Erkens, Malzahn, and Hoppe (2014). This again shows the importance of having a mix of modelling approaches and analysis techniques "at hand" to gain better insight and understanding of the determinants of learning and knowledge building communities.

The overarching claim and hope is to increase the awareness of the richness and variety of computational methods in the LA community, and thus to lay the ground for more synergy between LA and computationally inspired research.

REFERENCES

Bannert, M., Reimann, P., & Sonnenberg, C. (2014). Process mining techniques for analysing patterns and strategies in students' self-regulated learning. *Metacognition and Learning*, *9*(2), 161–185.

Batagelj, V. (2003). Efficient algorithms for citation network analysis. arXiv:cs/0309023v1.

Blei, D. M. (2012). Introduction to probabilistic topic models. *Communications of the ACM*, *55*(4), 77–84.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895.

Carley, K. M., Columbus, D., & Landwehr, P. (2013). AutoMap user's guide 2013. Technical Report CMU-ISR-13-105. Pittsburgh, PA: Carnegie Mellon University, Institute for Software Research. www.casos.cs.cmu.edu/publications/papers/CMU-ISR-13-105.pdf.

Daems, O., Erkens, M., Malzahn, N., & Hoppe, H. U. (2014). Using content analysis and domain ontologies to check learners' understanding of science concepts. *Journal of Computers in Education*, *1*(2–3), 113–131.

De Laat, M., Lally, V., Lipponen, L., & Simons, R. J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning*, *2*(1), 87–103.

Elkina, M., Fortenbacher, A., & Merceron, A. (2013). The learning analytics application LeMo: Rationals and first results. *International Journal of Computing*, *12*(3), 226–234.

Farooq, U., Schank, P., Harris, A., Fusco, J., & Schlager, M. (2007). Sustaining a community computing infrastructure for online teacher professional development: A case study of designing Tapped In. *Computer Supported Cooperative Work*, *16*(4–5), 397–429.

Ferguson, R., & Shum, S. B. (2012). Social learning analytics: Five approaches. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 23–33). New York: ACM.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*, *486*(3), 75–174.

Fournier-Viger, P., Gomariz, A., Gueniche, T., Soltani, A., Wu, C. W., & Tseng, V. S. (2014). SPMF: A Java open-source pattern mining library. *Journal of Machine Learning Research*, *15*(1), 3389–3393.

Fox, A. (2013). From MOOCs to SPOCs. *Communications of the ACM*, *56*(12), 38–40.

Halatchliyski, I., Oeberst, A., Bientzle, M., Bokhorst, F., & van Aalst, J. (2012). Unraveling idea development in discourse trajectories. *Proceedings of the 10th International Conference of the Learning Sciences* (ICLS '12), 2–6 July 2012, Sydney, Australia (vol. 2, pp. 162–166). International Society of the Learning Sciences.

Halatchliyski, I., Hecking, T., Göhnert, T., & Hoppe, H. U. (2014). Analyzing the path of ideas and activity of contributors in an open learning community. *Journal of Learning Analytics*, *1*(2), 72–93.

Harrer, A., Malzahn, N., Zeini, S., & Hoppe, H. U. (2007). Combining social network analysis with semantic relations to support the evolution of a scientific community. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of the 7th International Conference on Computer-Supported Collaborative Learning* (CSCL 2007), 16–21 July 2007, New Brunswick, NJ, USA (pp. 270–279). International Society of the Learning Sciences.

Harrer, A., Martínez-Monés, A., & Dimitracopoulou, A. (2009). Users' data: Collaborative and social analysis. In N. Balacheff, S. Ludvigsen, T. de Jong, A. Lazonder, & S. Barnes (eds.), *Technology-enhanced learning: Principles and products* (pp. 175–193). Springer Netherlands.

Haythornthwaite, C. (2001). Exploring multiplexity: Social network structures in a computer-supported distance learning class. *The Information Society*, 17(3), 211–226.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 29(1), 90–102.

Hecking, T., Ziebarth, S., & Hoppe, H. U. (2014). Analysis of dynamic resource access patterns in online courses. *Journal of Learning Analytics*, 1(3), 34–60.

Hoppe, H. U., Erkens, M., Clough, G., Daems, O. & Adams. A. (2013). Using Network Text Analysis to characterise teachers' and students' conceptualisations in science domains. In R. K. Vatrapu, P. Reimann, W. Halb, & S. Bull (Eds.), *Proceedings of the 2nd International Workshop on Teaching Analytics* (IWTA 2013), 8 April 2013, Leuven, Belgium. http://ceur-ws.org/Vol-985/paper7.pdf

Hoppe, H. U., Göhnert, T., Steinert, L., & Charles, C. (2014). A web-based tool for communication flow analysis of online chats. *Proceedings of the Workshops at the LAK 2014 Conference co-located with the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA. http://ceur-ws.org/Vol-1137.

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, 11(1), 39–63.

Hung, J. (2012). Trends of e-learning research from 2000 to 2008: Use of text mining and bibliometrics. *British Journal of Educational Technology*, 43(1), 5–16.

Krempel, L. (2005). *Visualisierung komplexer Strukturen – Grundlagen der Darstellung mehr-dimensionaler Netzwerke.* Frankfurt, Germany: Campus-Verlag.

Lehmann, S., Schwartz, M., & Hansen, L. K. (2008). Biclique communities. *Physical Review E*, 78(1). doi:10.1103/PhysRevE.78.016108

Mühlenbrock, M., & Hoppe, H. U. (1999). Computer supported interaction analysis of group problem solving. *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning* (CSCL '99), 12–15 December 1999, Palo Alto, California (pp. 398–405). International Society of the Learning Sciences.

Palla, G., Barabasi, A. L., & Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446(7136), 664–667.

Palla, G., Derenyi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043), 814–818.

Poquet, O., & Dawson, S. (2016). Untangling MOOC learner networks. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 208–212). New York: ACM.

Reffay, C., & Chanier, T. (2003). How social network analysis can help to measure cohesion in collaborative distance-learning. In B. Wasson, S. Ludvigsen, & H. U. Hoppe (Eds.), *Proceedings of the Conference on Computer Support for Collaborative Learning: Designing for Change in Networked Environments* (CSCL 2003), 14–18 June 2003, Bergen, Norway (pp. 343–352). International Society of the Learning Sciences.

Rosé, C., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237–271.

Sherin, B. (2013). A computational study of commonsense science: An exploration in the automated analysis of clinical interview data. *Journal of the Learning Sciences*, 22(4), 600–638.

Strijbos, J. W., Martens, R. L., Prins, F. J., & Jochems, W. M. (2006). Content analysis: What are they talking about? *Computers & Education*, 46(1), 29–48.

Suthers, D. D., Dwyer, N., Medina, R., & Vatraou, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer-Supported Collaborative Learning*, 5(1), 5–42.

Suthers, D. D., & Desiato, C. (2012). Exposing chat features through analysis of uptake between contributions. *Proceedings of the 45th Hawaii International Conference on System Sciences* (HICSS-45), 4–7 January 2012, Maui, HI, USA (pp. 3368–3377). IEEE Computer Society.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.

Wise, A. F., Cui, Y., & Vytasek, J. (2016). Bringing order to chaos in MOOC discussion forums with content-related thread identification. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 188–197). New York: ACM.

Wu, T., Khan, F. M., Fisher, T. A., Shuler, L. A., & Pottenger, W. M. (2005). Posting act tagging using transformation-based learning. In T. Y. Lin, S. Ohsuga, C.-J. Liau, X. Hu, & S. Tsumoto (Eds.), *Foundations of data mining and knowledge discovery* (pp. 321–331). Springer Berlin Heidelberg.

Zeini, S., Göhnert, T., Hecking, T., Krempel, L., & Hoppe, H. U. (2014). The impact of measurement time on subgroup detection in online communities. In F. Can, T. Özyer, & F. Polat (Eds.), *State of the art applications of social network analysis* (pp. 249–268). Springer International Publishing.

Ziebarth, S., Neubaum G., Kyewski E., Krämer N., Hoppe H. U., & Hecking T. (2015). Resource usage in online courses: Analyzing learner's active and passive participation patterns. *Proceedings of the 11th International Conference on Computer Supported Collaborative Learning* (CSCL 2015), 7–11 June 2015, Gothenburg, Sweden (pp. 395–402). International Society of the Learning Sciences.

Zumbach, J., Mühlenbrock, M., Jansen, M., Reimann, P., & Hoppe, H. U. (2002). Multi-dimensional tracking in virtual learning teams: An exploratory study. *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community* (CSCL 2002), 7–11 January 2002, Boulder, CO, USA (pp. 650–651). International Society of the Learning Sciences.

Chapter 3: Measurement and its Uses in Learning Analytics

Yoav Bergner

Learning Analytics Research Network, New York University, USA
DOI: 10.18608/hla17.003

ABSTRACT

Psychological measurement is a process for making warranted claims about states of mind. As such, it typically comprises the following: defining a construct; specifying a measurement model and (developing) a reliable instrument; analyzing and accounting for various sources of error (including operator error); and framing a valid argument for particular uses of the outcome. Measurement of latent variables is, after all, a noisy endeavor that can nevertheless have high-stakes consequences for individuals and groups. This chapter is intended to serve as an introduction to educational and psychological measurement for practitioners in learning analytics and educational data mining. It is organized thematically rather than historically, from more conceptual material about constructs, instruments, and sources of measurement error toward increasing technical detail about particular measurement models and their uses. Some of the philosophical differences between explanatory and predictive modelling are explored toward the end.

Keywords: Measurement, latent variable models, model fit

Knowing what students know and – given the increased attention to affective measures – how they feel is the basis for many conversations about learning. Measuring a student's knowledge, skills, attitudes/aptitudes/abilities (KSAs), and/or emotions is, however, less straightforward than measuring his or her height or weight. Psychological measurement is a noisy endeavor that can have high-stakes consequences, such as assignment to a special program (advanced or remedial), admission to a university, employment, hospitalization, or incarceration. Even small errors of measurement at the individual level can have large consequences when results are aggregated for groups (Kane, 2010). Sensitivity to these consequences has emerged over a century of methodology research enshrined in the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 2014). Insofar as measurement may be used in learning analytics and educational data mining for the purposes of understanding and optimizing learning and learning environments (Siemens & Baker, 2012), what are the tolerances for errors of measurement? After all, it has been argued that "harnessing the digital ocean" of data could ultimately replace the need for separate assessments (Behrens & DiCerbo, 2014). In the meantime, at minimum, one would like to avoid *misunderstanding* learning or *diminishing* learner experiences.

WHAT IS MEASUREMENT? PHILOSOPHY AND BASIC IDEAS

Discussions of psychological measurement often begin by drawing contrasts with physical measurement (for example, Armstrong, 1967; Borsboom, 2008; DeVellis, 2003; Lord & Novick, 1968; Maul, Irribarra, & Wilson, 2016; Michell, 1999; Sijtsma, 2011). A number of important facets of psychological measurement are raised in the process, namely its instrumentation or operationalization, the repeatability or precision of measurements, sources of error, and the interpretation of the measure itself. It can be said that psychological measurement comprises the following: defining a construct; specifying a measurement model and (developing) a reliable instrument; analyzing and accounting for various sources of error (including operator error); and framing a valid argument for particular uses of the outcome.

Constructs

Do psychological constructs really exist? In what sense can we really know a student's state of mind? We say that variables like physical length of an object are directly observed, or manifest, whereas a person's mental states or psychological traits are only indirectly observed, or latent. The term construct is used interchangeably with *latent variable*, while *trait* is used to imply a construct that is stable over time (Lord & Novick, 1968). In fact, even physical measurement is indirectly instrumented. Although we can perceive length directly through our senses, the *measurement* of length involves a process of comparison with a reference object or instrument, such as a tape measure. The tape measure provides a scale, such as inches or centimeters, which formalizes comparisons of length. For example, we can quantify the difference in two lengths by subtracting one measurement from the other.

In the first half of the twentieth century, efforts to settle philosophical issues of measurement led Bridgman (1927) and others to operationalism, wherein physical concepts like length, mass, and intensity are understood to be "synonymous with" the operations used to measure them. That is, length is understood as the outcome of a (possibly hypothetical) length measurement procedure. This idea can be carried over to psychological constructs, such as math ability and extraversion, by equating the constructs to scores on instruments used to measure them. Math ability is then equivalent to a score on a math test, and extraversion is a score on a Likert-item questionnaire. This positivist attitude is reflected in Stevens' definition of measurement as, "the assignment of numerals to objects or events according to rules" (1946, p. 677). The operationalist view of constructs was highly influential in the past, but it has been rejected for a host of reasons (Maul, Irribarra, & Wilson, 2016; Michell, 1999), notably that operationalism forces a redefinition of the construct for every instrument that exists to measure it.

If an operationalist interpretation is rejected, it appears to leave open epistemological and ontological questions about latent variables. Mislevy (2009, 2012) articulates a constructivist-realist position, namely that we can talk *as if* a construct exists without a commitment to strict realism by committing to model-based reasoning. Model-based reasoning means accepting a simplified representation of a system – for example, a construct-mediated relationship between persons and responses – that captures salient aspects (e.g., patterns) and allows us to explain or predict phenomena (Mislevy, 2009; we return to the explanatory/predictive distinction later in this chapter). As George Box famously said, "all models are wrong, but some are useful" (Box, 1979). The challenge remains to come up with useful models or, in terms of Stevens' definition, useful measurement rules.

Physical theories tend to be few in number and more comprehensive, whereas psychological theories are numerous and narrowly defined (DeVellis, 2003). Since constructs are invented things, there is no empirical limit to their number. It is possible to talk about a construct in the absence of a measurement *instrument*, but a measurement instrument is always designed to measure something. Therefore, we can infer an extremely partial list of constructs relevant to learning analytics from the instruments already developed to measure them. Examples include intelligence (e.g., the Stanford–Binet Intelligence Scale), scholastic aptitude (e.g., that SAT test), academic achievement (numerous examples include both large-scale tests and course exams), personality (e.g., the "big five" factor model; Digman, 1990), achievement-goal orientation (e.g., Midgley et al., 2000), achievement emotions (Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011), grit (Duckworth, Peterson, Matthews, & Kelly, 2007), self-theories of intelligence and fixed/growth mindset (Dweck, 2000; Yeager & Dweck, 2012), intrinsic motivation (Deci & Ryan, 1985; Guay, Vallerand, & Blanchard, 2000), self-regulated learning and self-efficacy (e.g., Pintrich & De Groot, 1990), learning power (Buckingham Shum & Deakin Crick, 2012; Crick, Broadfoot, & Claxton, 2004), and crowd-sourced learning ability (Milligan & Griffin, 2016).

Several of the constructs listed above are multidimensional, that is they include multiple factors. The value of combining versus separating out related constructs is a subject of debate (Edwards, 2001; Schwartz, 2007).

Measurement Instruments

Psychological measurement instruments are typically called tests or questionnaires (also surveys and inventories) and are made up of items or indicators. The word test is more often used for constructs like intelligence, cognitive ability, and psychomotor skills, wherein the subject, or examinee, is instructed to try to maximize his or her performance (Sijtsma, 2011). Questionnaire respondents, by contrast, are asked to respond honestly about their thoughts, feelings, and behaviours. (Response bias can blur this distinction, as we shall describe when we come to validity). Note that this description of how subjects are expected to interact with instruments reveals the rudiments of a measurement *model*. We assume that the more able test taker will obtain a higher score on an ability test and that the more anxious subject will obtain a higher score on an anxiety questionnaire.

Sometimes the term measurement scale is used interchangeably with the instrument (DeVellis, 2003). Scale implies that the test or questionnaire has been scored. Binary items that have correct and incorrect answers and yes/no questions are usually scored dichotomously with values in {0, 1}. Likert scale, rating scale, and visual-analogue scales (Luria, 1975) are other item types that can take discrete or continuous numerical values. Adding up the scores of individual items into a sum score (also, raw score) is one procedure for scoring an instrument, but it is not the only or necessarily the best procedure (Lord & Novick, 1968; Millsap, 2012). Weighted sum scores and item response theory (IRT; Baker & Kim, 2004) offer a range of alternatives.

The use of tests and questionnaires is a matter of both efficiency and standardization, compared with the alternative of observing people in real life and waiting for them to spontaneously express thoughts or exhibit the behaviours of interest (Sijtsma, 2011). In learning analytics, efficient collection of data is usually not the problem, but the lack of standardization can make it challenging to account for measurement error.

Source of Error in Measurements

We know from experience that psychological measurements are not as consistently repeatable as physical measurements. We also know that people's responses to an instrument may not faithfully reflect their abilities, attitudes, or other constructs of interest. Statistical models allow us to think of items, indicators, or tests as random samples of a latent variable. The latent variable can be a random variable, or it can be fixed, as in true score theory (Lord & Novick, 1968). Either way, the measurement samples will have error resulting from the inherent non-repeatability, which is sometimes called random error and is unbiased (in the sense of having an expectation value of zero over some distribution of repeated measures). There can also be systematic error, which is biased.

More precise or formal statements about error arise when we adopt a measurement framework or model. For example, in true score theory and factor analysis we can reason in terms of parallel tests or equivalent forms to derive estimates of an instrument's reliability. Measurement error can also be defined as any variance in the data not attributed to the construct, as explained by the model (AERA, APA, & NCME, 2014). We will revisit the sources of error after we flesh out our discussion of measurement models.

Reliability

Reliability is attributed to an instrument and is a measure of the consistency of scores (AERA, APA, & NCME, 2014), specifically the proportion of the total variance in scores attributed to the latent variable (DeVellis, 2003). It can be sample-dependent (in true score theory) and model-dependent (in more complicated models). The word is sometimes used to mean a particular reliability coefficient, most commonly Cronbach's (1951) alpha, α, which ranges from [0,1]. However, the term reliability is also used in the sense of test-retest reliability, which is actually a correlation, and inter-rater reliability (e.g., Cohen's kappa, κ; Cohen, 1968). Practitioners sometimes lean uncritically on guidelines for acceptable values of α, such as .70 as a lower bound (Cortina, 1993), to decide that scales are good enough to use. But it should be noted that statistical power improves with higher values of α (DeVellis, 2003). Thus, effort in improving the reliability of a scale can often outweigh the benefit of recruiting larger samples.

Validity

Validity is the foremost topic in the *Standards*, whose first chapter begins, "Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests … It is incorrect to use the unqualified phrase 'the validity of the test'" (p. 11). Substituting the broader term "measure" for the narrower "test," it should be self-evident that validity is of paramount importance to learning analytics. There is a palpable focus in the *Standards* on shaping the language used in validation arguments, an approach also evident in Messick's (1995) influential reworking of Cronbach and Meehl (1955) (see also Kane, 2001). Types of evidence about validity (rather than "types of validity") include evidence about response processes, evidence about the internal structure of the instrument, convergent and discriminant evidence, criterion references (including predictive criteria), and evidence of generalizability.

We referred earlier in this chapter to the assumption that responses to questionnaires correspond to honest thoughts and feelings. However, there is extensive literature on types of response bias, from acquiescence bias (yea-saying; Messick & Jackson, 1961) to social desirability bias (also, faking good; Nederhof, 1985) to bias from extreme and moderate types of responders (i.e., people who tend to choose extreme ends of Likert-scales) (Bachman & O'Malley, 1984). Although more often documented for questionnaires and surveys about sensitive topics such as willingness to cheat, sexual fantasies, or attitudes about race, self-tuning or censoring of responses can also happen on educational tests, such as the force concept inventory (FCI; Hestenes, Wells, & Swackhamer, 1992) used to assess Newtonian thinking. Mazur (2007) reported a student specifically asking, "How should I answer these questions? According to what you taught us, or by the way I think about these things?'" Finally, intentional rapid guessing behaviour can be thought of as a form of response bias (Wise & Kong, 2005). It should be clear that all of these sources of response bias challenge the uncritical interpretation of scale scores.

Measurement Models

The rubber meets the road in the technical details of measurement models. A measurement model is a formal mathematical relationship between a latent variable or set of variables and an observable variable or set of variables. A fully statistical measurement model may specify a distribution for the latent variable(s), a distribution for the observed variable(s), and a functional relationship between them. The latent variables are often understood as *causally explaining* the observations, which are subject to errors. Variances and covariances of random variables are described, explicitly or implicitly, in the model. Models make assumptions, for example the assumption of monotonicity (or, stricter, linearity) of the relationship between the construct and the measure or the assumption of zero covariance between error terms of unique items. If the assumptions of a model are violated, inferences made using the model may be wrong (Lord & Novick, 1968).

Since categorical and continuous variables involve different statistical methods, types of measurement models are sometimes classified into families according to the type of latent and observed variables, as shown in Table 3.1. This classification is not exhaustive, as hybrid models exist as well as generalized frameworks (Skrondal & Rabe-Hesketh, 2004) in which these model families become special cases. Growth models are extensions of measurement models to repeated measures and can apply to both continuous and categorical latent variables (e.g., Meredith & Tisak, 1990; Rabiner, 1989; Raudenbush & Bryk, 2002).

Table 3.1. Families of Latent Variable Models

Latent/Observed	Observed continuous	Observed categorical
Latent continuous	Factor models (Bollen, 1989; Mulaik, 2009)	Item response models (Lord & Novick, 1968; Baker & Kim, 2004)
Latent categorical	Latent mixture models (McLachlan & Peel, 2004)	Latent class models (Goodman, 2002)

SPECIFIC USES OF MEASUREMENT MODELS IN LEARNING ANALYTICS

We mentioned previously that psychological and educational measurement is applied for a variety of purposes including classification, diagnosis, ranking, placement, and certification of individuals as well as corresponding inferences about groups. Work in learning analytics and educational data mining also explores the complex web of relationships between psychological scales, behaviour, and performance in digital learning environments (Tempelaar, Rienties, & Giesbers, 2015). The purpose of this section is to provide a bit more depth about models and their uses in learning analytics and educational data mining. All topics are not treated equally, reflecting both space constraints and selection bias.

Factor Analysis

Factor analysis (Mulaik, 2009) models the correlations among observed variables through a linear relationship to a set of latent variables known as factors. The original one-factor model is Spearman's (1904) model of general intelligence g, used to explain correlations between scores on unrelated subject tests. True score theory, also known as classical test theory (Lord & Novick, 1968), can be derived as a special case of a single factor model in which all of the item factor loadings are the same. Thurstone (1947) developed the multiple (seven) factors model of intelligence.

Factor analysis is commonly divided into two enterprises. Exploratory factor analysis (EFA) is used to determine the number of latent factors from data without strong theoretical assumptions and is commonly part of scale development. However, EFA requires a number of important methodological decisions which, if made poorly, can lead to problematic results (Fabrigar, Wegener, MacCallum, & Strahan, 1999). In particular, Fabrigar et al. (1999) caution against confusing EFA with principal components analysis (PCA), a dimensionality reduction technique, which can result in erroneous conclusions about true factor structure. Confirmatory factor analysis (CFA) is a complementary set of techniques to test a theoretically proposed factor model by examining residuals between expected and observed correlations. Thus, CFA can be used to reject a model. CFA, along with path analysis and latent growth models, is subsumed by structural equation modelling (SEM; Bollen, 1989; Kline, 2010). Confirmatory factor analysis is not the same thing as running EFA multiple times with different population samples, although the case has been made for doing the latter (DeVellis, 2003).

Some learning analytics research is directly concerned with scale development and its integration with data gathered from learning management systems (e.g., Buckingham Shum & Deakin Crick, 2012; Milligan & Griffin, 2016). Other work focuses on associations between existing scales and outcome measures, such as the relationship between achievement emotions (Pekrun et al., 2011) and decisions regarding face-to-face and online instruction (Tempelaar, Niculescu, Rienties, Giesbers, & Gijselaers, 2012) or between motivational measures and completion of a massive open online course (Wang & Baker, 2015). When adapting an instrument or, especially, part of an instrument for new purposes, practitioners should be mindful of whether these new uses merit new validation arguments.

Latent Class and Latent Mixture Models

Dedic, Rosenfeld, and Lasry (2010) used latent class analysis to understand the distribution of physics misconceptions based on students' wrong answers on a physics concept test. Data came from administrations both before and at the end of a physics course (pre- and post-test). The authors identified an apparent progression from Aristotelian to Newtonian thinking through discrete classes of dominance fallacies. A widely used method for topic modelling of documents, latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003; see also several chapters in this volume) is a latent mixture model. Mixed membership models (Erosheva, Fienberg, & Lafferty, 2004) further generalize latent mixtures by allowing "fuzzy" or weighted assignments of an individual to multiple classes. The Gaussian mixture model forms the basis for model-based cluster analysis (Fraley & Raftery, 1998) applied to performance trajectories of MOOC learners (Bergner, Kerr, & Pritchard, 2015). It should be noted that not all clustering algorithms, however, are latent mixture models.

Item Response Theory (IRT)

Item response theory distinguished itself in the historical development of testing theory by modelling individual person-item interactions rather than total test scores, as in classical test theory. Conceptually, the purpose of IRT is "to describe the items by item parameters and the examinees by examinee parameters in such a way that we can predict probabilistically the response of any examinee to any item, even if similar examinees have never taken similar items before" (Lord, 1980, p. 11). A sample item characteristic curve (ICC) or, equivalently, item response function (IRF) for a binary item (e.g., correct/incorrect, agree/disagree, et cetera) is shown in Figure 3.1.

The salient characteristics of Figure 3.1 are as follows:

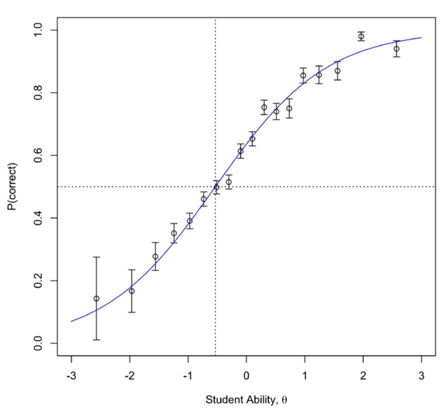


Figure 3.1. A sample item characteristic curve (ICC). Dotted lines indicate the P = 0.5 intercept.

1. The trait (e.g., ability) is quantified as a continuous random variable and is represented by θ on the horizontal axis. The variable is standardized to have a mean of zero and a variance of 1 in the population of interest. More of the trait, corresponding to a higher value of θ, is expected to increase the probability P of a positive (or correct) response. This is the *monotonicity assumption*. An observed violation of monotonicity means that that the fundamental person-item relationship is wrong, and including the item in a test would lead to bad fit and unreliable inferences.

2. Two ways of interpreting these curves were described by Holland (1990). In the stochastic subject interpretation, one literally imagines this curve as applying to an individual whose performance is inherently unpredictable. To paraphrase Holland, the stochastic subject explanation is intuitive, but not wholly satisfactory; we do not have a mechanistic explanation for the stochastic nature of the subject. In the random sampling interpretation, on the other hand, this curve makes sense as applied to a sample population of examinees. For example, among examinees within a certain ability range, some proportion will answer correctly. The points and error bars in the figure reflect this observation.[1]

3. The value of θ for which P = 0.5 is a reference intercept, which for a cognitive ability test item is called the difficulty. Note that difficulty is *ipso facto* on the same scale as ability, and so it makes sense to talk about the difference between a person's ability and the difficulty of an item.

4. The form of the probability link is commonly parametric with respect to the trait θ_i of individual i and a (set of) item parameters β_j, for item j,

$$P_{ij} = P(X_{ij} = 1|\theta_i, \beta_j) = f(\theta_i, \beta_j), \tag{1}$$

as in the case of the Rasch model (a single difficulty parameter) or of the two-parameter logistic (2PL) model. The 2PL model is shown in Figure 3.2; the fit to data is visibly good, and a G^2 goodness-of-fit test confirms as much. It should be noted that non-parametric IRT methods exist (Sijtsma, 1998).

When a person responds to several items in a measurement instrument, the idea is to combine the response information to make posterior estimates of the trait. For the likelihood of a response vector to factor into a product of individual item-level probabilities, the responses must be otherwise independent, conditional on the trait. This *conditional independence assumption*

[1] For the stochastic subject, these sample values would have to represent a set of identical trials by the same subject with no memory of the other trials. Although this seems odd in a cognitive test item, it is plausible in a psychomotor context. See Spray (1997).

may require the introduction of additional factors that explain inter-item dependence (e.g., Rijmen, 2010).

Evidence that IRT has some traction in education outside of high-stakes testing applications can be found in physics education research applications to the force concept inventory (FCI; Hestenes et al., 1992) and the mechanics baseline test (MBT; Hestenes & Wells, 1992). While these instruments have been in use for twenty-five years, item response model analyses started to appear more recently (Morris et al., 2006; Wang & Bao, 2010). Model-data fit for the FCI were generally acceptable. Cardamone et al. (2011), however, discovered two malfunctioning items in the MBT by inspecting the item response functions. An example is shown in Figure 3.2.

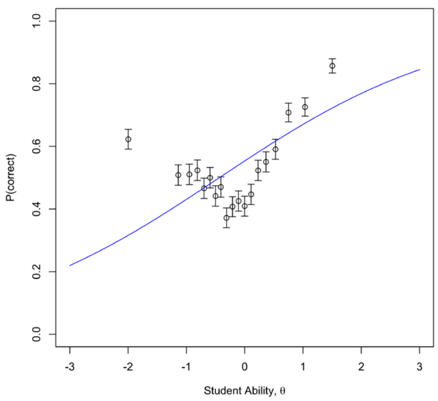


Figure 3.2. A poorly fitting item from the mechanics baseline test (MBT).

Something is fishy if low-ability students are more likely to answer an item correctly than average-ability students. Upon closer inspection, it was discovered that ambiguous wording of this test item allowed students holding a common misconception to misread the question and coincidentally choose the correct response for the wrong reason. In this case, two wrongs did make a right.

Following exploratory factor analyses of the FCI that identified multiple dimensions (Ding & Beichner, 2009; Scott, Schumayer, & Gray, 2012), a variation of multidimensional IRT was applied to the MBT (Bergner, Rayyan, Seaton, & Pritchard, 2013). Item response theory models have also been extended to the inherently sequential process behind multiple attempts to answer (answer-until-correct), an affordance which is common in online homework (Attali, 2011; Bergner, Colvin, & Pritchard, 2015; Culpepper, 2014).

Growth Models

Growth models apply any time a latent trait is changing systematically between measurements. They can be applied to changing attitudes (e.g., George, 2000), but we focus here on application to cognitive ability domains. There is an extensive literature in educational data mining on student models for intelligent problem-solving tutors, which are distinguished from curriculum sequencing tutors (Desmarais & Baker, 2011).

In cognitive tutors for mathematics (Anderson, Corbett, Koedinger, & Pelletier, 1995), sequences of practice items are designed to support mastery learning of fine-grained knowledge components (also, skills or productions), according to a cognitive model. Two approaches for modelling growth towards mastery in data from these systems are Bayesian knowledge tracing (BKT; Corbett & Anderson, 1995) and the additive factors models (AFM; Cen, Koedinger, & Junker, 2008; Draney, Pirolli, & Wilson, 1995). Learning curves analysis (Käser, Koedinger, & Gross, 2014; Martin, Mitrovic, Mathan, & Koedinger, 2010) has also been used to check for discrepancies between data and the cognitive model underlying the tutor.

According to the "law of practice" (Newell & Rosenbloom, 1981), the aggregate error rate T as a function of practice opportunity n should decay according to a power law $T=B_n^{-\alpha}$, where B and a are empirically determined. Bad fit between data and model, for example using r-squared measures, may motivate improvements to knowledge mapping. This may be seen as an analogue to the item analysis in Figure 3.2, where a faulty item is detected. In this case, however, the assignment of a sequence of items to a knowledge component is seen as faulty.

In BKT, the latent variable is mastery of a procedural knowledge component and is binary-valued, $M \in \{0, 1\}$. The probability link between mastery and correctness $X \in \{0, 1\}$ on any given opportunity is a 2x2 conditional probability table, but by analogy with Eq. (1), it can be written in terms of guess (g) and slip (s) parameters as,

$$P(X = 1|M) = (1 - s)^M g^{(1-M)} \quad (2)$$

Importantly, the attempts are not viewed as independent. Rather, the key idea in BKT is that students begin with some prior probability of mastery and move towards mastery (they learn) on each practice opportunity according to the rule,

$$P(M_n) = P(M_{n-1}) + \tau(1 - P(M_{n-1})) \quad (3)$$

Here τ is a growth parameter. Recently, van de Sande (2013) demonstrated that BKT implies an exponential rather than a power law relationship between practice attempts and error rates. This would make BKT a mis-specified model for data that satisfy a power law of practice. The additive factors model, by contrast, is designed to fit the power law of practice paradigm. Käser et al. (2014) showed that prediction accuracy of BKT is often indistinguishable from AFM. Regarding fit of the latter, they noted systematic bias in aggregate residuals analyses.

AFM has been referred to as an extension of IRT (Koedinger, McLaughlin, & Stamper, 2012), and indeed the relation to the linear logistic test model (LLTM; Fischer, 1973) was clear in the progenitor of this model (Draney et al., 1995). However, in passing to its current form, the model was changed in a critical way. The LLTM is a Rasch-type IRT model in which the difficulty of an item is decomposed as a sum over potential properties of the item. Writing the Rasch model as,

$$logit(P_{ij}) = ln(P_{ij}/(1-P_{ij})) = \theta_i - \beta_j, \tag{4}$$

the difficulty β_j of item j is further decomposed,

$$\beta_j = c_j + \textstyle\sum_k w_{jk}\alpha_k, \tag{5}$$

where α_k are difficulties of "basic" operations (Fischer's term) and the indicators w_{ik} are 0 or 1 depending on whether these operations are required in item j. If all items use the same operations, the model clearly reduces to the Rasch model with a simple offset,

$$\beta_j = c_j + \alpha. \tag{6}$$

Although the model of Draney et al. (1995) contained an item-level difficulty parameter, in AFM only the difficulties of the component skills are retained. In addition, a practice term is introduced,[2]

$$\beta_j^{AFM} = \textstyle\sum_k w_{jk}\alpha_k - \sum_k w_{jk}\gamma_k T_{ik}, \tag{7}$$

where γ_k is a growth parameter and T_{ik} is a count of the previous practice attempts of learner i on skill k. If a sequence of practice problems all involve the same skills, which is common for tutor applications, then for each sequence, this parameter reduces to,

$$\beta_j^{AFM} = \alpha - \gamma T_i. \tag{8}$$

Importantly, this is not a property of the item at all, as is clear from the subscripts on the right hand side, which depend only on the learner. By dropping the c_j parameter in Equations (7)–(8), the AFM has actually become a fixed effect growth model.

From a modelling perspective, it is not surprising that the item-level difficulty parameter was removed, as keeping both difficulty and growth parameters creates a problem for identifiability. A model is identifiable if its parameters can be unambiguously learned given sufficient data. However, for students working on a fixed sequence of items, the increased success rate due to learning/growth can be attributed to decreasing item difficulty. The two effects cannot be distinguished unless item difficulties have been separately calibrated under conditions where there is no growth.

Cognitive Diagnostic Models

A seminal study of mixed-number subtraction using cognitive task analysis led Tatsuoka (1983) to develop the Q-matrix method and a model for diagnosing specific sub-skills (e.g., converting a whole number to a fraction) in an educational test. The Q-matrix is a discrete mapping of items to requisite sub-skills and is traditionally specified in the assessment model. Cognitive diagnostic models have since been considerably generalized (Rupp & Templin, 2008; von Davier, 2005), and efforts to learn the Q-matrix from data have appeared in educational data mining research (Barnes, 2005; Desmarais, 2012; Koedinger et al., 2012).

SOURCES OF ERROR, REVISITED

Having explored some of the measurement models involved in studying motivation, emotion, and cognition, it is worth revisiting the important subject of error. Practitioners should be mindful that additional sources of error could be introduced by using models with the wrong parameters, by using the wrong models, or by using the models wrongly.

The use of a model may depend on parameters whose estimation is itself subject to error. These uncertainties should be acknowledged, but they are not necessarily serious if the model is consistent as a *data-generating model* for the observed data. That is, we think of the statistical model as a stochastic process that can be used to generate (also, sample or simulate) data (Breiman, 2001). For example, we can simulate data from coin flips using a Bernoulli process, even if we are unsure about whether the real coin is fair. In principle, our parameter for the probability of heads in our model can be improved with more data from the real coin. This is different from the case when the model itself, either in terms of the latent variables or the link functions, is inconsistent with the true generating model. The second case is called model mis-specification (White, 1996). Goodness-of-fit tests evaluate the consistency between the observed data and the generating model to retain or reject the model (White, 1996; Haberman, 2009; Ames & Penfield, 2015).

EXPLANATION AND PREDICTION

Predictive modelling is one of the most prominent methodological approaches in educational data mining (Baker & Siemens, 2014; Baker & Yacef, 2009). Measurement theory, by contrast, is decidedly explanatory, as are most of the statistical methods traditionally used in the social sciences (Breiman, 2001; Shmueli,

[2] One sign convention from Cen et al. (2008) has been changed to make the model consistent with the usual Rasch model, with a difficulty rather than an easiness parameter.

2010). While an explanatory model can be used to make predictions – and an error-free explanatory model would make perfect predictions – a predictive model is not necessarily explanatory. Breiman (2001) expressed the distinction in terms of two cultures: the data modelling culture (98% of statistics, informally according to Breiman) and the algorithmic modelling culture (the 2%, in which Breiman included himself).[3] Shmueli (2010) contrasted the entire design process for statistical modelling when viewed from either a prediction or an explanation lens. The interpretability or non-interpretability of predictors in a complex prediction model is only one aspect of the distinction (see also Liu & Koedinger, this volume). The different viewpoints fundamentally inform how researchers handle error and uncertainty.

The predictive view is expressed, for example, in a recent best paper from the educational data mining conference. The authors assert that, "the only way to determine if model assumptions are correct is to construct an alternative model that makes different assumptions and to determine whether the alternative outperforms [out-predicts] BKT" (Khajah, Lindsey, & Mozer, 2016, p. 95, editorial note added). Strictly speaking, model prediction performance is not a way to determine if model assumptions are violated. By contrast, both informal checks and formal tests for goodness-of-fit have been discussed above. However, the quote is a reflection of the algorithmic modelling culture in which models are validated by predictive accuracy (Breiman, 2001). More problematically, it carries a presumption that predictive power points to the truer model. In fact, it is explanatory power that plays this role. Put in terms of variance components, "in explanatory modelling the focus is on minimizing bias to obtain the most accurate representation of the underlying theory. In contrast, predictive modelling seeks to minimize the combination of bias and variance, occasionally sacrificing theoretical accuracy for improved empirical precision" (Shmueli, 2010, p. 293). It should be emphasized that explanatory power and predictive power do not always point in the same direction. Indeed, Hagerty and Srinivasan (1991) proved that, in noisy circumstances, under-specified multiple regression models can have more predictive power than the correctly specified (true) model.

Suthers and Verbert (2013) have described learning analytics as a "middle space" between learning science and analytics. Perhaps it may also be thought of as occupying a methodological middle space between explanatory and predictive approaches. In that case, the field may benefit from understanding the nuances of both perspectives.

FURTHER READING

Psychological measurement is almost as old as psychology itself and as old as statistics. Authoritative, technical, and somewhat encyclopedic sources are the anthology of psychometrics in the *Handbook of Statistics* series (Rao & Sinharay, 2006) and the "bible" of *Educational Measurement*, now in its fourth edition (Brennan, 2006). Educational measurement volumes and the *Standards* (AERA, APA, & NCME, 2014) tend to emphasize testing, where specific issues are reliability, validity, generalizability, comparability, and fairness. DeVellis' (2003) concise volume on scale development is a non-technical introduction to psychological measurement and omits topics specific to large-scale testing, such as linking scores from parallel test forms.

[3] Breiman uses the term *information* in place of *explanation* and in contrast to *prediction*.

REFERENCES

AERA, APA, & NCME (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Ames, A. J., & Penfield, R. D. (2015). An NCME instructional module on polytomous item response theory models. *Educational Measurement: Issues and Practice*, 34(3), 39–48. doi:10.1111/emip.12023

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences*, 4(2), 167–207.

Armstrong, J. S. (1967). Derivation of theory by means of factor analysis or Tom Swift and his electric factor analysis machine. *The American Statistician*, 21, 17–21.

Attali, Y. (2011). Immediate feedback and opportunity to revise answers: Application of a graded response IRT model. *Applied Psychological Measurement*, 35(6), 472–479.

Baker, F. B., & Kim, S.-H. (Eds.). (2004). *Item response theory: Parameter estimation techniques*. Boca Raton, FL: CRC Press.

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In R. Sawyer (Ed), *The Cambridge handbook of the learning sciences* (pp. 253–272). Cambridge University Press.

Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.

Barnes, T. (2005). The Q-matrix method: Mining student response data for knowledge. In the Technical Report (WS-05-02) of the AAAI-05 Workshop on Educational Data Mining.

Behrens, J. T., & DiCerbo, K. E. (2014). Harnessing the currents of the digital ocean. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 39–60). New York: Springer.

Bachman, J. G., & O'Malley, P.M. (1984). Yea-saying, nay-saying, and going to extremes: Black-white differences in response styles. *Public Opinion Quarterly*, 48, 491–509.

Bergner, Y., Colvin, K., & Pritchard, D. E. (2015). Estimation of ability from homework items when there are missing and/or multiple attempts. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 118–125). New York: ACM.

Bergner, Y., Kerr, D., & Pritchard, D. E. (2015). Methodological challenges in the analysis of MOOC data for exploring the relationship between discussion forum views and learning outcomes. In O. C. Santos et al. (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 234–241). International Educational Data Mining Society.

Bergner, Y., Rayyan, S., Seaton, D., & Pritchard, D. E. (2013). Multidimensional student skills with collaborative filtering. *AIP Conference Proceedings*, 1513(1), 74–77. doi:10.1063/1.4789655

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan.), 993–1022.

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Borsboom, D. (2008). Latent variable theory. *Measurement: Interdisciplinary Research & Perspective*, 6(1–2), 25–53. http://doi.org/10.1080/15366360802035497

Box, G. E. (1979). Robustness in the strategy of scientific model building. *Robustness in Statistics*, 1, 201–236.

Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199–215. http://doi.org/10.2307/2676681

Brennan, R. L. (Ed.). (2006). *Educational measurement*. Praeger Publishers.

Bridgman, P. W. (1927). *The logic of modern physics*. New York: Macmillan.

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modeling and learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 92–101). New York: ACM.

Cardamone, C. N., Abbott, J. E., Rayyan, S., Seaton, D. T., Pawl, A., & Pritchard, D. E. (2011). Item response theory analysis of the mechanics baseline test. *Proceedings of the 2011 Physics Education Research Conference* (PERC 2011), 3–4 August 2011, Omaha, NE, USA (pp. 135–138). doi:10.1063/1.3680012

Cen, H., Koedinger, K. R., & Junker, B. (2008). Comparing two IRT models for conjunctive skills. In B. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (ITS 2008), 23–27 June 2008, Montreal, PQ, Canada (pp. 796–798). Springer.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70(4), 213–220.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4, 253–278.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98.

Crick, R. D., Broadfoot, P., & Claxton, G. (2004). Developing an effective lifelong learning inventory: The ELLI project. *Assessment in Education: Principles, Policy & Practice*, 11(3), 247–272.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52(4), 281–302.

Culpepper, S. A. (2014). If at first you don't succeed, try, try again: Applications of sequential IRT models to cognitive assessments. *Applied Psychological Measurement*, 38(8), 632–644. doi:10.1177/0146621614536464

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum.

Dedic, H., Rosenfield, S., & Lasry, N. (2010). Are all wrong FCI answers equivalent? *AIP Conference Proceedings*, 1289, 125–128. doi.org/10.1063/1.3515177

Desmarais, M. C. (2012). Mapping question items to skills with non-negative matrix factorization. *ACM SIGKDD Explorations Newsletter*, 13(2), 30–36.

Desmarais, M. C., & Baker, R. S. (2011). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1–2), 9–38. doi:10.1007/s11257-011-9106-8

DeVellis, R. F. (2003). *Scale development: Theory and applications*. Applied Social Research Methods Series (Vol. 26). Thousand Oaks, CA: Sage Publications.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, 41(1), 417–440.

Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics: Physics Education Research*, 5(2), 1–17. doi:10.1103/PhysRevSTPER.5.020103

Draney, K., Pirolli, P., & Wilson, M. R. (1995). A measurement model for a complex cognitive skill. In P. Nichols, S. Chipman, & R. Brennan (Eds.), *Cognitively diagnostic assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 9, 1087–1101.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality and development*. Philadelphia, PA: Taylor & Francis.

Edwards, J. R. (2001). Multidimensional constructs in organizational behavior research: An integrative analytical framework. *Organizational Research Methods*, 4(2), 144–192.

Erosheva, E., Fienberg, S., & Lafferty, J. (2004). Mixed-membership models of scientific publications. *Proceedings of the National Academy of Sciences*, 101(suppl 1), 5220–5227.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272.

Fischer, G. H. (1973). The linear logistic test model as an instrument in educational research. *Acta Psychologica*, 37(6), 359–374.

Fraley, C., & Raftery, A. E. (1998). How many clusters? Which clustering method? Answers via model-based cluster analysis. *The Computer Journal*, 41(8), 578–588.

George, R. (2000). Measuring change in students' attitudes toward science over time: An application of latent variable growth modeling. *Journal of Science Education and Technology*, 9(3), 213–225.

Goodman, L. (2002) Latent class analysis: The empirical study of latent types, latent variables, and latent structures. In J. A. Hagenaars & A. L. McCutcheon (Eds.), *Applied latent class analysis* (pp. 3–55). Cambridge, UK: Cambridge University Press.

Guay, F., Vallerand, R. J., & Blanchard, C. (2000). On the assessment of situational intrinsic and extrinsic motivation: The situational motivation scale (SIMS). *Motivation and Emotion*, 24(3), 175–213.

Haberman, S. J. (2009). Use of generalized residuals to examine goodness of fit of item response models. *ETS Research Report RR-09-15*.

Hagerty, M. R., & Srinivasan, V. (1991). Comparing the predictive powers of alternative multiple regression models. *Psychometrika*, 56(1), 77–85.

Hestenes, D., & Wells, M. (1992). A mechanics baseline test. *The Physics Teacher*, 30(3), 159–166.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30(3), 141. doi:10.1119/1.2343497

Holland, P. W. (1990). On the sampling theory roundations of item response theory models. *Psychometrika*, 55(4), 577–601. http://doi.org/10.1007/BF02294609

Kane, M. T. (2001). Current concerns in validity theory. *Journal of Educational Measurement*, 38(4), 319–342.

Kane, M. (2010). Errors of measurement, theory, and public policy. William H. Angoff Memorial Lecture Series. *Educational Testing Service*. https://www.ets.org/Media/Research/pdf/PICANG12.pdf

Käser, T., Koedinger, K. R., & Gross, M. (2014). Different parameters – same prediction: An analysis of learning curves. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July 2013, Memphis, TN, USA (pp. 52–59). International Educational Data Mining Society/Springer.

Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing? In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (EDM2016), 29 June–2 July 2016, Raleigh, NC, USA (pp. 94–101). International Educational Data Mining Society.

Kline, R. B. (2010). *Principles and practice of structural equation modeling*. New York: Guilford.

Koedinger, K. R., McLaughlin, E. A., & Stamper, J. (2012). Automated student model improvement. In K. Yacef et al. (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (EDM2012), 19–21 June 2012, Chania, Greece. International Educational Data Mining Society. http://www.learnlab.org/research/wiki/images/e/e1/KoedingerMcLaughlinStamperEDM12.pdf

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Luria, R. E. (1975). The validity and reliability of the visual analogue mood scale. *Journal of Psychiatric Research*, 12(1), 51–57.

Martin, B., Mitrovic, T., Mathan, S., & Koedinger, K. R. (2010). Evaluating and improving adaptive educational systems with learning curves. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 21, 249–283.

Maul, A., Irribarra, D. T., & Wilson, M. (2016). On the philosophical foundations of psychological measurement. *Measurement*, 79, 311–320. http://doi.org/10.1016/j.measurement.2015.11.001

Mazur, E. (2007). Confessions of a converted lecturer. https://www.math.upenn.edu/~pemantle/active-papers/Mazurpubs_605.pdf

McLachlan, G., & Peel, D. (2004). *Finite mixture models*. John Wiley & Sons.

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, 55(1), 107–122.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.

Messick, S., & Jackson, D. (1961). Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin*, 58(4), 299–304

Michell, J. (1999). *Measurement in psychology: A critical history of a methodological concept* (Vol. 53). Cambridge University Press.

Midgley, C., Maehr, M. L., Hruda, L., Anderman, E. M., Anderman, L., Freeman, K. E., et al. (2000). *Manual for the patterns of adaptive learning scales (PALS)*. Ann Arbor, MI: University of Michigan.

Milligan, S. K., & Griffin, P. (2016). Understanding learning and learning design in MOOCs: A measurement-based interpretation. *Journal of Learning Analytics*, 3(2), 88–115.

Millsap, R. E. (2012). *Statistical approaches to measurement invariance*. Routledge.

Mislevy, R. J. (2009). Validity from the perspective of model-based reasoning. In R. L. Lissitz (Ed.), *The concept of validity: Revisions, new directions and applications* (pp. 83–108). Charlotte, NC: Information Age Publishing.

Mislevy, R. J. (2012). Four metaphors we need to understand assessment. Draft paper commissioned by the Gordon Commission. http://www.gordoncommission.com/rsc/pdfs/mislevy_four_metaphors_understand_assessment.pdf

Morris, G. A., Branum-Martin, L., Harshman, N., Baker, S. D., Mazur, E., Dutta, S., … McCauley, V. (2006). Testing the test: Item response curves and test quality. *American Journal of Physics*, 74(5), 449. doi:10.1119/1.2174053

Mulaik, S. A. (2009). *Foundations of factor analysis*. Boca Raton, FL: CRC Press.

Nederhof, A. J. (1985). Methods of coping with social desirability bias: A review. *European Journal of Social Psychology*, 15(3), 263–280. http://doi.org/10.1002/ejsp.2420150303

Newell, A., & Rosenbloom, P. S. (1981). Mechanisms of skill acquisition and the law of practice. *Cognitive Skills and their Acquisition*, 6, 1–55.

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The achievement emotions questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1), 36–48. http://doi.org/10.1016/j.cedpsych.2010.10.002

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.

Rao, C. R., & Sinharay, S. (Eds.). (2006). *Handbook of statistics 26: Psychometrics*. Elsevier. doi:10.1016/S0169-7161(06)26037-1

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.

Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47(3), 361–372. doi:10.1111/j.1745-3984.2010.00118.x

Rupp, A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research & Perspective*, 6(4), 219–262. doi:10.1080/15366360802490866

Schwartz, S. (2007). The structure of identity consolidation: Multiple correlated constructs or one superordinate construct? *Identity*, 7(1), 27–49.

Scott, T. F., Schumayer, D., & Gray, A. R. (2012). Exploratory factor analysis of a force concept inventory data set. *Physical Review Special Topics: Physics Education Research*, 8(2). doi:10.1103/PhysRevSTPER.8.020105

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. http://doi.org/10.1214/10-STS330

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 252–254). New York: ACM.

Sijtsma, K. (2011). Introduction to the measurement of psychological attributes. *Measurement*, 44(7), 1209–1219. doi:10.1016/j.measurement.2011.03.019

Sijtsma, K. (1998). Methodology review: Nonparametric IRT approaches to the analysis of dichotomous item scores. *Applied Psychological Measurement*, 22(1), 3–31. doi:10.1177/01466216980221001

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton, FL: Chapman & Hall/CRC Press.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–292.

Spray, J. A. (1997). Multiple-attempt, single-item response models. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 209–220). New York: Springer.

Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677–680.

Suthers, D., & Verbert, K. (2013). Learning analytics as a middle space. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 1–4). New York: ACM.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345–354.

Tempelaar, D. T., Niculescu, A., Rienties, B., Giesbers, B., & Gijselaers, W. H. (2012). How achievement emotions impact students' decisions for online learning, and what precedes those emotions. *Internet and Higher Education*, 15(3), 161–169. doi:10.1016/j.iheduc.2011.10.003

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47, 157–167. doi:10.1016/j.chb.2014.05.038

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago, IL: University of Chicago Press.

van de Sande, B. (2013). Properties of the Bayesian knowledge tracing model. *Journal of Educational Data Mining*, 5(2), 1–10.

von Davier, M. (2005). A general diagnostic model applied to language testing data. *The British Journal of Mathematical and Statistical Psychology*, 61(Pt 2), 287–307. doi:10.1348/000711007X193957

Wang, Y., & Baker, R. S. (2015). Content or platform: Why do students complete MOOCs? *Journal of Online Learning and Teaching*, 11(1), 17.

Wang, J., & Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics*, 78(10), 1064. doi:10.1119/1.3443565

White, H. (1996). *Estimation, inference and specification analysis* (No. 22). Cambridge University Press.

Wise, S., & Kong, X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18(2), 163–183.

Yeager, D. S., & Dweck, C. S. (2012). Mindsets that promote resilience: When students believe that personal characteristics can be developed. *Educational Psychologist*, 47(4), 302-314.

Chapter 4: Ethics and Learning Analytics: Charting the (Un)Charted

Paul Prinsloo,[1] Sharon Slade[2]

[1] Department of Business Management, University of South Africa, South Africa
[2] Faculty of Business and Law, The Open University, United Kingdom

DOI: 10.18608/hla17.004

ABSTRACT

As the field of learning analytics matures, and discourses surrounding the scope, definition, challenges, and opportunities of learning analytics become more nuanced, there is benefit both in reviewing how far we have come in considering associated ethical issues and in looking ahead. This chapter provides an overview of how our own thinking has developed and maps our journey against broader developments in the field. Against a backdrop of technological advances and increasing concerns around pervasive surveillance and the role and unintended consequences of algorithms, the development of research in learning analytics as an ethical and moral practice provides a rich picture of fears and realities. More importantly, we begin to see ethics and privacy as crucial enablers within learning analytics. The chapter briefly locates ethics in learning analytics in the broader context of the forces shaping higher education and the roles of data and evidence before tracking our personal research journey, highlighting current work in the field, and concluding by mapping future issues for consideration.

Keywords: Ethics, artificial intelligence, data, big data, students

In 2011, the New Media Consortium's Horizon Report (NMC, 2011) pointed to the increasing importance of learning analytics as an *emerging* technology, which has since developed from a mid-range technology or trend to one to be realized within a "one year or less" time-frame (NMC 2016, p. 38). Though there are clear linkages between learning analytics and the more established field of educational data mining, there are also important distinctions regarding, inter alia, automation, aims, origins, techniques, and methods (Siemens & Baker, 2012). As the field of learning analytics has developed as a distinct field of research and practice (see van Barneveld, Arnold, & Campbell, 2012), so too thinking around ethical issues has slowly moved in from the margins. Slade and Prinsloo (2013) established one of the earliest frameworks developed with a focus on ethics in learning analytics. Since then, the number of authors publishing in this sub-field has significantly increased, resulting in a growing number of frameworks, codes of practices, taxonomies, and guidelines (Gašević, Dawson & Jovanović, 2016).

In the wider context of public concerns surrounding increasing surveillance and the (un)warranted collection, analysis, and use of personal data, "fears and realities are often indistinguishably mixed up, leading to an atmosphere of uncertainty among potential beneficiaries" (Drachsler & Greller, 2016, p. 89). Gašević et al. (2016) also suggest that further challenges remain "to be addressed in order to further aid uptake and integration into educational practice," and see ethics and privacy as important enablers to the field of learning analytics "rather than barriers" (p. 2).

We briefly situate set the context for considering the ethical implications of learning analytics, before mapping our personal research journey in the field. We then consider recent developments and conclude by flagging a selection of issues that continue to require broader and more critical engagement.

SETTING THE CONTEXT: WHY ETHICS IS RELEVANT

There is some consensus that the future of learning will be digital, distributed, and data-driven such that education "enables quality of life and meaningful employment through (a) exceptional quality research; (b) sophisticated data collection and; (c) advanced machine learning and human learning analysis/ support" (Siemens, 2016, Slide 2). Although considerations of the ethical implications of learning analytics were initially on the margins of the field, the prominence of ethics has come a long way and is increasingly foregrounded (Gašević et al., 2016). In a context where much is to be said for the potential *economic* benefits (for both students and the institution) of more successful learning experiences resulting from increased data harvesting, we should not ignore the possibilities of "data-proxy-induced hardship… when the detail obtained from the data-proxy comes to disadvantage its embodied referent in some way" (Smith, 2016, p. 16; also see Ruggiero, 2016; Strauss, 2016b; and Watters, 2016).

Ethical implications around the collection, analysis, and use of student data should take cognizance of the potentially conflicting interests and claims of a range of stakeholders, such as students and institutions. Views on the benefits, risks, and potential for harm resulting from the collection, analysis, and use of student data will depend on the interests and perceptions of the particular stakeholder. In this chapter, we hope to provide insight into the different positionalities, claims, and interests of primarily students and institutions.

ESTABLISHING ETHICAL PRINCIPLES: HOW FAR HAVE WE COME?

Although now becoming more established, ethics and the need to question *how* student data is used and under what conditions was very much a marginal issue in the early years of the field. The first attempts to explore wider issues around learning analytics were presented at LAK '12 in Vancouver. The large majority of sessions at this conference remained focused on developmental work. There was some mention of stakeholder perceptions of the ways in which student data could be used, notably Drachsler and Greller (2012), though their paper suggested that surveyed stakeholder considerations were largely focused on privacy and not considered particularly contentious. A further paper (Prinsloo, Slade, & Galpin, 2012) touched upon the need to consider the impacts of *all* stakeholders on students' learning journeys in order to increase the success of students' learning. The notion of "thirdspace" provided a useful heuristic to map the challenges and opportunities, but also the paradoxes of learning analytics and its potential impact on student success and retention. At the same conference, an exploratory workshop (Slade & Galpin, 2012) built upon early work by Campbell, DeBlois, and Oblinger (2007) aiming to consider and expand upon a number of assumed relevant ethical issues from different stakeholder perspectives.

Work in 2013 moved onto an examination of existing institutional policy frameworks that set out the purposes for how data would be used and protected (Prinsloo & Slade, 2013). The growing advent of learning analytics had seen uses of student data expanding rapidly. In general, policies relating to institutional use of student data had not kept pace, nor taken account of the growing need to recognize ethical concerns, focusing mainly on data governance, data security, and privacy issues. The review identified clear gaps and the insufficiency of existing policy.

Taking a sociocritical perspective on the use of learning analytics, Slade and Prinsloo (2013) considered a number of issues affecting the scope and definition of the ethical use of learning analytics. A range of ethical issues was grouped within three broad, overlapping categories, namely:

- The location and interpretation of data
- Informed consent, privacy, and the de-identification of data
- The management, classification, and storage of data

Slade and Prinsloo (2013) proposed a framework based on the following six principles:

1. Learning analytics as moral practice – focusing not only on what is effective, but on what is appropriate and morally necessary
2. Students as agents – to be engaged as collaborators and not as mere recipients of interventions and services
3. Student identity and performance as temporal dynamic constructs – recognizing that analytics provides a snapshot view of a learner at a particular time and context
4. Student success as a complex, multidimensional phenomenon
5. Transparency as important – regarding the purposes for which data will be used, under what conditions, access to data, and the protection of an individual's identity
6. That higher education cannot afford not to use data

These principles offer a useful starting position, but ought sensibly to be supported by consideration of a number of practical considerations, such as the development of a thorough understanding of who benefits (and under what conditions); establishment of institutional positions on consent, de-identification and opting out; issues around vulnerability and harm (e.g., inadvertent labelling); systems of redress (for both student and institution); data collection, analyses, access, and storage (e.g., security issues and avoiding perpetuation of bias); and governance and resource allocation (including clarity around the key drivers for "success" (and what success means), existing constraints, and the conditions that must be met).

This latter aspect of resource allocation was carefully explored in a later paper considering the concept of educational triage (Prinsloo & Slade, 2014a). Although learning analytics offers *theoretical* opportunities for HEIs (higher education institutions) to proactively identify and support students at risk of failing or dropping out, they do so in a context whereby resources are (increasingly) limited. The challenge then is *where best* to direct support resources and on what basis that decision is made. The concept of educational triage as a means of directing support toward students most likely to "survive" requires careful consideration of a number of related and complex issues, such as the balance between respecting student autonomy and, at the same time, ensuring the long-term sustainability of the institution; the notion of beneficence (to always act in the student's best interest); the need for non-maleficence (inflicting the least harm possible to reach a beneficial outcome); and maintaining a sense of distributive justice (understanding that demographic characteristics have and do impact support provided and assumptions made, and the need to recognize and address this).

An increasing awareness of learning analytics as a means of *doing something to* the student without that student necessarily knowing triggered further exploration of issues around surveillance, student privacy and institutional accountability (Prinsloo & Slade, 2014b). The resulting discussion challenged assumptions around learning analytics as a producer of accurate, objective, fully complete pictures of student learning, and also reviewed the potentially unequal relationship between institution and student. In considering existing frameworks regarding the use and analysis of personal data, the study suggested six elements that could form a basis for a *student-centred* learning analytics:

1. The use of aggregated, non-personalized data is essential in delivering effective and appropriate teaching and learning, but students should be able to make informed opt in/out decisions
2. Students should have full(er) knowledge of which data is collected and how it is used
3. Students should ensure that their personal data records are complete and up to date
4. The surveillance of activities and the harvesting of data must not harm student progress
5. Algorithmic output should be subject to (potential) human review, and corrected if needed
6. Learning analytics essentially provides context and time-specific, provisional, incomplete pictures of students, and algorithms should be frequently reviewed and validated

Issues around surveillance and the need to recognize students as active agents in the use of their own data was explicitly addressed within the development of The Open University (OU; 2014) policy on the ethical use of student data for learning analytics. As part of the stakeholder consultation, a representative group of 50 students explored their understanding of the ways in which data is used to support students in completing their study goals over a three-week period. A study of responses (Slade & Prinsloo, 2014) found that students appeared largely unaware of the extent to which data was already actively collected and used, and they raised a number of concerns. The major concern related to the potential to actively consent (or not), with a majority of students expressing a wish for a right to opt out. This direct involvement of student voices in shaping a policy dealing with the ethics of learning analytics offered unique insight into the ways in which students regard their data – as a valuable entity to be carefully protected and even more carefully applied. Given that the sample in Slade and Prinsloo (2014) may not be fully representative of the total population, the outcomes cannot be generalized across institutional and geopolitical contexts.

In response to this growing awareness of student concern, Prinsloo and Slade (2015) questioned whether our assumptions and understanding of issues surrounding student attitudes to privacy may be influenced by both the apparent ease with which the public appear to share the detail of their lives and by our largely paternalistic institutional cultures. The study explored issues around consent and the seemingly simple choice to allow students to opt-in or opt-out of having their data tracked. As a foundation for the discussion, the terms and conditions of three massive open online course (MOOC) providers were reviewed to establish information given to users regarding the uses of their data. This extended into a discussion of how HEIs can move toward an approach that engages and more fully informs students of the implications of learning analytics on their personal data. A similar theme was pursued in Prinsloo and Slade (2016a). This paper challenged the tendency for many HEIs to adopt an authoritarian approach to student data. Despite the rapid growth in the deployment of learning analytics, few HEIs have regulatory frameworks in place and/or are fully transparent regarding the scope of student data collected, analyzed, used, and shared. Student vulnerability was explored in the nexus between realizing the potential of learning analytics; the fiduciary duty of HEIs in the context of their asymmetrical information and power relations with students; and the complexities surrounding student agency in learning analytics. The aim was to consider ways in which student vulnerability may be addressed, increasing student agency, and empowering them as active participants in learning analytics – moving from quantified data objects to qualified and qualifying selves (see also Prinsloo & Slade, 2016b).

RECENT DEVELOPMENTS IN ETHICAL FRAMEWORKS

It is broadly accepted that the increasing value of data as a sharable commodity with an increasing exchange value has overtaken our legal and traditional ethical frameworks (Zhang, 2016). "Deep economic pressures are driving the intensification of connection and monitoring online" (Couldry, 2016, par. 13) and "What's needed is more collective reflection on the costs of capitalism's new data relations for our very possibilities of ethical life" (Couldry, 2016, par. 35). As such, there have been attempts in different geopolitical and institutional contexts to grapple with the ethical implications of learning analytics. Sclater, Peasgood, and Mullan (2016), for example, review practices within higher education in the United States, Australia, and the United Kingdom. They summarize their findings by indicating that learning analytics makes significant contributions for 1) quality assurance and quality improvement; 2) boosting retention rates; 3) assessing and acting upon differential outcomes among the student population; and 4) the development and introduction of adaptive learning. The report acknowledges the many opportunities, but also highlights threats such as "ethical and data privacy issues, 'over-analysis' and the lack of generalizability of the results, possibilities for misclassification of patterns, and contradictory findings" (p. 16). In their review, the one institutional example of a policy level bid to address the ethical concerns in learning analytics is that of The Open University (UK). In 2014, the OU published a "Policy on ethical use of student data for learning analytics" delimiting the nature and scope of data collected and analyzed and an explicit specification of data that will *not* be collected and used for learning analytics. The policy establishes the following eight principles (p. 6):

1. Learning analytics is an ethical practice that should align with core organizational principles, such as open entry to undergraduate level study.
2. The OU has a responsibility to all stakeholders to use and extract meaning from student data for the benefit of students where feasible.
3. Students should not be wholly defined by their visible data or our interpretation of that data. [This principle furthermore warns against stereotyping students and acknowledges those individuals who do not fit into typical patterns. The principle also makes it clear that members of staff may not have access to the full data set, which can seriously impact the reliability of the analysis.]
4. The purpose and the boundaries regarding the use of learning analytics should be well defined and visible.
5. The University is transparent regarding data collection, and will provide students with the opportunity to update their own data at regular intervals.
6. Students should be engaged as active agents in the implementation of learning analytics (e.g., informed consent, personalized learning paths, interventions).
7. Modelling and interventions based on analysis of data should be sound and free from bias.
8. Adoption of learning analytics within the OU requires broad acceptance of the values and benefits (organizational culture) and the development of appropriate skills across the organization.

As one of the first institutional responses to the ethical implications in the collection, analysis, and use of student data, this policy and its principles attempted to map uncharted territory. Of specific interest is the definition of "informed consent" as referring to "the process whereby the student is made aware of the purposes to which some or all of their data may be used for learning analytics and provides consent. Informed consent applies at the point of reservation or registration on to a module or qualification" (Open University, 2014, p. 3). The policy does not address the possibility of students who prefer to opt out of the collection, analysis, and use of their data (as discussed by Engelfriet, Manderveld, & Jeunink, 2015; Sclater, 2015; also see Shacklock, 2016).

In a recent overview of learning analytics practices in the Australian context, Dawson, Gašević, and Rogers (2016) report that the "relative silence afforded to ethics across the studies is significant" (p. 3) and that this "does not reflect the seriousness with which the sector should consider these issues" (p. 33). The report suggests that "It is likely that the higher education sector has not been ready for such a conversation previously, although it is argued that as institutions are maturing, ethical considerations take on a heightened salience" (p. 33).

Also in the Australian context, Welsh and McKinney (2015) position the need for a Code of Practice in learning analytics in the context of the "relative immaturity of the discipline with institutions, practitioners and technology vendors still figuring out what works and finding the boundaries of 'acceptable' practice" and the real potential for abuse/misuse and discrimination (p. 588). Of particular importance is the commitment that "The University will not engage in Learning Analytics practices that use data sources: (a) not directly related to learning and teaching; and/or (b) where users may not reasonably expect such data collection by the University to occur" (p. 590). Student data will only be used in the context of the original purpose for which the data in question was collected; its use can continue under the following conditions:

> Explicit informed consent is gathered from those who are the subject of measurement. Where informed consent means that: (a) clear and accurate information is provided about what data is or may be collected, why and how it is collected, how it is stored and how it is used; and (b) agreement is freely given to the practice(s) described. (p. 590)

The above principles should be read in conjunction with two remaining principles regarding how collected data should be used to enhance teaching and learning and to give students "greater control over and responsibility for their learning" (p. 591); and one of transparency and informed participation. For a full discussion, see Welsh and McKinney (2015).

Drachsler and Greller (2016) provide a broad overview of ethics, privacy, and respective legal frameworks, and highlight challenges such as the real possibility of exploitation in light of the asymmetrical power relationship between data gatherer and data object, issues of ownership, anonymity and data security, privacy and data identity, as well as transparency and trust. They present a checklist (DELICATE©) to ensure that learning analytics proceeds in an acceptable and compliant way "to overcome the fears connected to data aggregation and processing policies" (p. 96).

Sclater (2015) proposes a (draft) taxonomy of ethical, legal, and logistical issues in learning analytics with an overview of how a range of stakeholders, such as senior management, the analytics committee, data scientists, educational researchers, IT, and students are impacted and have responsibility in learning analytics. The draft covers a wide range of issues including, inter alia, consent; identity; potential impacts of opting out; the asymmetrical relationship between the institution and students; (boundaries around) the permissible uses of student data; transparency; data included (and excluded) from use; and student autonomy, amongst others. See Sclater (2015) for a full list of ethical concerns.

In the Dutch higher education context, Engelfriet et al. (2015) consider the implications of the Law for the Protection of Personal Information for learning analytics. These include the need for permission (and the responsibility arising from receiving consent) and the implications of the consensual agreement between a service provider and recipient that the provider may use any personal information needed for the provision of the service. The law distinguishes between *essential* information and "handy" information. Engelfriet et al. (2015) take a contentious view that, given that learning analytics is seen as an emerging practice, it may safely be regarded as collecting "handy" information, and so perhaps excluded from the need for consensual agreement between the institution and students. The authors suggest that these four principles should guide learning analytics:

- Personal information be used only in the context and purpose for which it was provided
- Subsequent use of such data should be reconcilable with the original context and purpose
- Data should be carefully collected and analyzed, and "sneaky" ("stiekeme" in Dutch) usage of analytics is not permissible; this appears to emphasize a need for transparency, student consultation, and buy in
- Data may only be collected when the purpose/use of the collected data is made explicitly clear

Engelfriet et al. (2015) explore student rights around the governance of their data, including the following:

- Easy access to collected information
- The right to correct wrong information (or interpretations arising from it)
- The right to remove irrelevant information

Of particular interest is an exploration of the ethical implications for algorithmic decision making and the authors flag examples that lead to potential conflict with Dutch law. The implication is that humans need to take responsibility for and have oversight of algorithmic decision making. Algorithms may, at most, signal particular behaviour for the attention of faculty or support staff. Further, students have a right to appeal decisions made based on analyses of their personal data. In cases where HEIs subcontract to software developers, the final responsibility and oversight remains securely with the institution and cannot be delegated (see Engelfriet et al., 2015).

SOME FUTURE CONSIDERATIONS

It falls outside the scope of this chapter to map current and future gaps in our understanding of the complexities and practicalities at the intersections between student data and advances in technology and methods of analysis. We would like to conclude, however, with some pointers for future consideration.

Given the mandate of higher education institutions to ensure effective, appropriate, cost-effective learning experiences and to support students to be successful, there is broad agreement that institutions have a right to collect and use student information. However, there is no easily agreed upon position around consent, that is, in allowing students to opt out of the collection, analysis, and use of their data. Student positions around consent may be influenced by issues not wholly logical or rational. The often-implicit calculation of benefits, costs, and risks will depend on a range of factors such as, inter alia, previous experiences, need, and perceived benefits (see, for example, Daniel Pink in O'Brien. 2010).

One recent example of opt out was led by the National Center for Fair and Open Testing in the US who encouraged students to refuse to take government-mandated standardized tests. Around 650,000 students opted out in the 2014–2015 school year (FairTest, n.d.), with the US Department of Education responding by threatening to withhold funding (Strauss, 2016a).

Further research is needed to explore potential conflicts between students' concerns, their right to opt-out, and the implications for the mandate of higher education to use student data to make interventions at an individual level. Central to this issue is the question of "who benefits?" (see Watters, 2016). Any consideration of the ethics around the collection, analysis, and use of student data (whether in learning analytics or in formal assessments) should also recognize the contesting claims and vested interests.

In the broader context of online research, Vitak, Shilton, and Ashktorab (2016) point to various challenges regarding ethical research practices in online contexts, such as the increasing and persisting concerns about re-identification: "researchers still struggle to balance research ethics considerations with the use of online datasets" (p. 1). Interestingly, their findings also show that many the researchers go beyond the Belmont principles (with the main emphasis on ensuring that outcomes outweigh potential harms caused by the research) by referring to "(1) transparency with participants, (2) ethical deliberation with colleagues, and (3) caution in sharing results" (par. 66).

There is also increasing concern balancing optimism around artificial intelligence (AI), machine learning, and big data. For example, the Executive Office of the President of the US released a report (Munoz, Smith, & Patil, 2016) that highlights benefits, but also addresses concerns regarding the potential harm inherent in the use of big data. The report recognizes that if "these technologies [algorithmic systems] are not implemented with care, they can also perpetuate, exacerbate, or mask harmful discrimination" (p. 5). It makes a number of suggestions relating to investment in research into the mitigation of algorithmic discrimination, encouraging the development and use of robust and transparent algorithms, algorithmic auditing, improvements in data science "fluency," and the roles of the government and private sector in setting codes of practice around data use.

Similarly, the UK Government recently released a "Data science ethical framework" (Cabinet Office, 2016) providing guidance on "ethical issues which sit outside the law" (p. 3). The framework explores issues such as the nature of the benefits of the collection, analysis, and use of personal data; the scope and nature of intrusion; the quality of the data and the automation of the decisions relating to the collected data; the risk of negative unintended consequences; whether the data objects agreed to the collection and analysis; the nature and scope of the oversight; and the security of the collected data. The framework also proposes a "Privacy Impact Assessment" requiring data scientists to clarify "tricky issues" (p. 6), such as reviewing the extent to which the benefits of the project outweigh the risks to privacy and negative unintended consequences; steps undertaken to minimize risks and ensure correct interpretation; and the extent to which the opinions of the data objects/public regarding the project were considered (see Cabinet Office, 2016).

In the context of the algorithmic turn in (higher) education, and the increasing blurring of the boundaries between broader developments in data and neuroscience, we need a critical approach to considering the ethical implications of learning analytics as we find our way through the myth, mess, and methods (Ziewitz, 2016) of student data. For example, Williamson (2016a) considers "educational data science as a *biopolitical strategy* focused on the evaluation and management of the corporeal, emotional and embrained lives of children" (p. 401, emphasis added). As such, we have to consider the basis and scope of authority of educational data scientists who have "increasing legitimate authority to produce systems of knowledge about children and to define them as subjects and objects of intervention" (Williamson, 2016a, p. 401). Learning analytics in future will be essentially based on and driven by algorithms and machine learning and we therefore have to consider how algorithms "reinforce, maintain, or even reshape visions of the social world, knowledge, and encounters with information" (Williamson, 2016b, p. 4). Accountability, transparency, and regulatory frameworks will be essential elements in the frameworks ensuring *ethical* learning analytics (see Prinsloo, 2016; Taneja, 2016).

While this chapter maps the progress in considering the ethical implications of the collection, analysis, and use of student data, it is clear that the potential for harm will not be addressed without further consideration of institutional processes to ensure accountability and transparency. As Willis, Slade, and Prinsloo (2016) indicate, learning analytics often falls outside the processes and oversight provided by institutional review boards (IRBs). It is not clear at this stage by whom and how the ethical implications of learning analytics will be assured.

(IN)CONCLUSIONS

Since the emergence of learning analytics in 2011, the field has not only matured, but also become more nuanced in increasingly considering the fears and realities of ethical implications in the collection, analysis, and use of student data. In this chapter, we provide an overview of how our own thinking has developed alongside broader developments in the field. Against a backdrop of technological advances and increasing concerns around pervasive surveillance, and a growing consensus that the future of higher education will be digital, distributed, and data-driven, this chapter maps how far the discourses surrounding the ethical implications of learning analytics have come, as well as some of the future considerations.

Each of the frameworks, code of practices, and conceptual mappings of the ethical implications in learning analytics discussed adds a further layer and a richer understanding of how we may move toward using student data-proxies to increase the effectiveness and appropriateness of teaching, learning, and student support strategies in economically viable and ethical ways. The practical implementation of that understanding remains largely incomplete, but still wholly pertinent.

ACKNOWLEDGEMENTS

We would like to acknowledge the comments, critical inputs, and support received from the editorial team and specially the reviewers of this chapter.

REFERENCES

Cabinet Office. (2016, 19 May). Data science ethical framework. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/524298/Data_science_ethics_framework_v1.0_for_publication__1_.pdf

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007, July/August) Academic analytics: A new tool, a new era. *EDUCAUSE Review*. http://net.educause.edu/ir/library/pdf/erm0742.pdf

Couldry, N. (2016, September 23). The price of connection: "Surveillance capitalism." [Web log post]. https://theconversation.com/the-price-of-connection-surveillance-capitalism-64124

Dawson, S., Gašević, D., & Rogers, T. (2016). Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement. Australian Government. http://he-analytics.com/wp-content/uploads/SP13_3249_Dawson_Report_2016-3.pdf

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics understandings and expectations from the stakeholders. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 120–129). New York: ACM.

Drachsler, H., & Greller, W. (2016, April). Privacy and analytics: It's a DELICATE issue – a checklist for trusted learning analytics. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 89–98). New York: ACM.

Engelfriet, A., Manderveld, J., & Jeunink, E. (2015). Learning analytics onder de Wet bescherming persoonsgegevens. SURFnet. https://www.surf.nl/binaries/content/assets/surf/nl/kennisbank/2015/surf_learning-analytics-onder-de-wet-wpb.pdf

FairTest. (n.d.). Just say no to the test. http://www.fairtest.org/get-involved/opting-out

Gašević, D., Dawson, S., & Jovanović, J. (2016). Ethics and privacy as enablers of learning analytics. *Journal of Learning Analytics*, 3(1), 1–4. http://dx.doi.org/10.18608/jla.2016.31.1

Munoz, C., Smith, M., & Patil, D. J. (2016). Big data: A report on algorithmic systems, opportunity, and civil rights. Executive Office of the President, USA. https://www.whitehouse.gov/sites/default/files/micro-sites/ostp/2016_0504_data_discrimination.pdf

NMC (New Media Consortium). (2011). The NMC Horizon Report. http://www.educause.edu/Resources/2011HorizonReport/223122

NMC (New Media Consortium). (2016). The NMC Horizon Report. http://www.nmc.org/publication/nmc-horizon-report-2016-higher-education-edition/

O'Brien, A. (2010, September 29). Predictably irrational: A conversation with best-selling author Dan Ariely. [Web log post]. http://www.learningfirst.org/predictably-irrational-conversation-best-selling-author-dan-ariely

Open University. (2014). Policy on ethical use of student data for learning analytics. http://www.open.ac.uk/students/charter/sites/www.open.ac.uk.students.charter/files/files/ecms/web-content/ethical-use-of-student-data-policy.pdf

Prinsloo, P. (2016, September 22). Fleeing from Frankenstein and meeting Kafka on the way: Algorithmic decision-making in higher education. Presentation at NUI, Galway. http://www.slideshare.net/prinsp/feeling-from-frankenstein-and-meeting-kafka-on-the-way-algorithmic-decisionmaking-in-higher-education

Prinsloo, P., Slade, S., & Galpin, F. (2012) Learning analytics: Challenges, paradoxes and opportunities for mega open distance learning institutions. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 130–133). New York: ACM.

Prinsloo, P., & Slade, S. (2013). An evaluation of policy frameworks for addressing ethical considerations in learning analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 240–244). New York: ACM.

Prinsloo, P., & Slade, S. (2014a). Educational triage in higher online education: Walking a moral tightrope. *International Review of Research in Open Distributed Learning* (IRRODL), 14(4), 306–331. http://www.irrodl.org/index.php/irrodl/article/view/1881

Prinsloo, P., & Slade, S. (2014b). Student privacy and institutional accountability in an age of surveillance. In M. E. Menon, D. G. Terkla, & P. Gibbs (Eds.), *Using data to improve higher education: Research, policy and practice* (pp. 197–214). Global Perspectives on Higher Education (29). Rotterdam: Sense Publishers.

Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 83–92). New York: ACM.

Prinsloo, P., & Slade, S. (2016a). Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182.

Prinsloo, P., & Slade, S. (2016b). Here be dragons: Mapping student responsibility in learning analytics. In M. Anderson & C. Gavan (Eds.), *Developing effective educational experiences through learning analytics* (pp. 170–188). Hershey, PA: IGI Global.

Ruggiero, D. (2016, May 18). What metrics don't tell us about the way students learn. *The Conversation*. http://theconversation.com/what-metrics-dont-tell-us-about-the-way-students-learn-59271

Sclater, N. (2015, March 3). Effective learning analytics. A taxonomy of ethical, legal and logistical issues in learning analytics v1.0. JISC. https://analytics.jiscinvolve.org/wp/2015/03/03/a-taxonomy-of-ethical-legal-and-logistical-issues-of-learning-analytics-v1-0/

Sclater, N., Peasgood, A., & Mullan, J. (2016). Learning analytics in higher education. A review of UK and international practice. JISC. https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education

Shacklock, X. (2016). From bricks to clicks: The potential of data and analytics in higher education. Higher Education Commission. http://www.policyconnect.org.uk/hec/sites/site_hec/files/report/419/fieldreportdownload/frombrickstoclicks-hecreportforweb.pdf

Siemens, G. (2016, April 28). Reflecting on learning analytics and SoLAR. [Web log post]. http://www.elearnspace.org/blog/2016/04/28/reflecting-on-learning-analytics-and-solar/

Siemens, G., & Baker, R. (2012, April). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 252–254). New York: ACM.

Slade, S., & Galpin, F. (2012) Learning analytics and higher education: Ethical perspectives (workshop). *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 16–17). New York: ACM.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, 57(1), 1509–1528.

Slade, S., & Prinsloo, P. (2014). Student perspectives on the use of their data: Between intrusion, surveillance and care. *Proceedings of the European Distance and E-Learning Network 2014 Research Workshop* (EDEN 2014), 27–28 October 2014, Oxford, UK (pp. 291–300).

Smith, G. J. (2016). Surveillance, data and embodiment: On the work of being watched. *Body & Society*, 1–32. doi:10.1177/1357034X15623622

Strauss, V. (2016a, January 28). U.S. Education Department threatens to sanction states over test opt-outs. The Washington Post. https://www.washingtonpost.com/news/answer-sheet/wp/2016/01/28/u-s-education-department-threatens-to-sanction-states-over-test-opt-outs/

Strauss, V. (2016b, May 9). "Big data" was supposed to fix education. It didn't. It's time for "small data." The Washington Post. https://www.washingtonpost.com/news/answer-sheet/wp/2016/05/09/big-data-was-supposed-to-fix-education-it-didnt-its-time-for-small-data/

Taneja, H. (2016, September 8). The need for algorithmic accountability. *TechCrunch*. https://techcrunch.com/2016/09/08/the-need-for-algorithmic-accountability/

van Barneveld, A., Arnold, K., & Campbell, J. (2012). Analytics in higher education: Establishing a common language. *EDUCAUSE Learning Initiative*, *1*, 1–11.

Vitak, J., Shilton, K., & Ashktorab, Z. (2016). Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community. *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '16), 27 February–2 March 2016, San Francisco, CA, USA. New York: ACM. https://terpconnect.umd.edu/~kshilton/pdf/VitaketalCSCWpreprint.pdf

Watters, A. (2016, May 7). Identity, power, and education's algorithms. [Web log post]. http://hackeducation.com/2016/05/07/identity-power-algorithms

Welsh, S., & McKinney, S. (2015). Clearing the fog: A learning analytics code of practice. In T. Reiners et al. (Eds.), Globally connected, digitally enabled. *Proceedings of the 32nd Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (ASCILITE 2015), 29 November–2 December 2015, Perth, Western Australia (pp. 588–592). http://research.moodle.net/80/

Williamson, B. (2016a). Coding the biodigital child: The biopolitics and pedagogic strategies of educational data science. *Pedagogy, Culture & Society*, 24(3), 401–416. doi:10.1080/14681366.2016.1175499

Williamson, B. (2016b). Computing brains: Learning algorithms and neurocomputation in the smart city. *Information, Communication & Society*, 20(1), 81–99. doi:10.1080/1369118X.2016.1181194

Willis, J., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: A typology derived from a cross-continental, cross-institutional perspective. Educational Technology Research and Development. doi:10.1007/s11423-016-9463-4

Zhang, S. (2016, May 20). Scientists are just as confused about the ethics of big data research as you. Wired. http://www.wired.com/2016/05/scientists-just-confused-ethics-big-data-research/

Ziewitz, M. (2016). Governing algorithms: Myth, mess, and methods. *Science, Technology & Human Values*, 41(1), 3–16.

SECTION 2

TECHNIQUES & APPROACHES

Chapter 5: Predictive Modelling in Teaching and Learning

Christopher Brooks[1], Craig Thompson[2]

[1] School of Information, University of Michigan, USA
[2] Department of Computer Science, University of Saskatchewan, Canada
DOI: 10.18608/hla17.005

ABSTRACT

This article describes the process, practice, and challenges of using predictive modelling in teaching and learning. In both the fields of educational data mining (EDM) and learning analytics (LA) predictive modelling has become a core practice of researchers, largely with a focus on predicting student success as operationalized by academic achievement. In this chapter, we provide a general overview of considerations when using predictive modelling, the steps that an educational data scientist must consider when engaging in the process, and a brief overview of the most popular techniques in the field.

Keywords: Predictive modeling, machine learning, educational data mining (EDM), feature selection, model evaluation

Predictive analytics are a group of techniques used to make inferences about uncertain future events. In the educational domain, one may be interested in predicting a measurement of learning (e.g., student academic success or skill acquisition), teaching (e.g., the impact of a given instructional style or specific instructor on an individual), or other proxy metrics of value for administrations (e.g., predictions of retention or course registration). Predictive analytics in education is a well-established area of research, and several commercial products now incorporate predictive analytics in the learning content management system (e.g., D2L,[1] Starfish Retention Solutions,[2] Ellucian,[3] and Blackboard[4]). Furthermore, specialized companies (e.g., Blue Canary,[5] Civitas Learning[6]) now provide predictive analytics consulting and products for higher education.

In this chapter, we introduce the terms and workflow related to predictive modelling, with a particular emphasis on how these techniques are being applied in teaching and learning. While a full review of the literature is beyond the scope of this chapter, we encourage readers to consider the conference proceedings and journals associated with the Society for Learning Analytics and Research (SoLAR) and the International Educational Data Mining Society (IEDMS) for more examples of applied educational predictive modelling.

First, it is important to distinguish predictive modelling from explanatory modelling.[7] In explanatory modelling, the goal is to use all available evidence to provide an explanation for a given outcome. For instance, observations of age, gender, and socioeconomic status of a learner population might be used in a regression model to explain how they contribute to a given student achievement result. The intent of these explanations is generally to be causal (versus correlative alone), though results presented using these approaches often eschew experimental studies and rely on theoretical interpretation to imply causation (as described well by Shmueli, 2010).

In predictive modelling, the purpose is to create a model that will predict the values (or class if the prediction does not deal with numeric data) of new data based on observations. Unlike explanatory modelling, predictive modelling is based on the assumption that a set of known data (referred to as *training instances* in data mining

[1] http://www.d2l.com/
[2] http://www.starfishsolutions.com/
[3] http://www.ellucian.com/
[4] http://www.blackboard.com/
[5] http://bluecanarydata.com/
[6] http://www.civitaslearning.com/

[7] Shmueli (2010) notes a third form of modelling, descriptive modelling, which is similar to explanatory modelling but in which there are no claims of causation. In the higher education literature, we would suggest that causation is often implied, and the majority of descriptive analyses are actually intended to be used as causal evidence to influence decision making.

literature) can be used to predict the value or class of new data based on observed variables (referred to as *features* in predictive modelling literature). Thus the principal difference between explanatory modelling and predictive modelling is with the application of the model to future events, where explanatory modelling does not aim to make any claims about the future, while predictive modelling does.

More casually, explanatory modelling and predictive modelling often have a number of pragmatic differences when applied to educational data. Explanatory modelling is a post-hoc and reflective activity aimed at generating an understanding of a phenomenon. Predictive modelling is an in situ activity intended to make systems responsive to changes in the underlying data. It is possible to apply both forms of modelling to technology in higher education. For instance, Lonn and Teasley (2014) describe a student-success system built on explanatory models, while Brooks, Thompson, and Teasley (2015) describe an approach based upon predictive modelling. While both methods intend to inform the design of intervention systems, the former does so by building software based on theory developed during the review of explanatory models by experts, while the latter does so using data collected from historical log files (in this case, clickstream data).

The largest methodological difference between the two modelling approaches is in how they address the issue of generalizability. In explanatory modelling, all of the data collected from a sample (e.g., students enrolled in a given course) is used to describe a population more generally (e.g., all students who could or might enroll in a given course). The issues related to generalizability are largely based on sampling techniques. Ensuring the sample represents the general population by reducing selection bias, often through random or stratified sampling, and determining the amount of power needed to ensure an appropriate sample, through an analysis of population size and levels of error the investigator is willing to accept. In a predictive model, a *hold out* dataset is used to evaluate the suitability of a model for prediction, and to protect against the overfitting of models to data being used for training. There are several different strategies for producing hold out datasets, including k-fold cross validation, leave-one-out cross validation, randomized subsampling, and application-specific strategies.

With these comparisons made, the remainder of this chapter will focus on how predictive modelling is being used in the domain of teaching and learning, and provide an overview of how researchers engage in the predictive modelling process.

PREDICTIVE MODELLING WORKFLOW

Problem Identification

In the domain of teaching and learning, predictive modelling tends to sit within a larger action-oriented educational policy and technology context, where institutions use these models to react to student needs in real-time. The intent of the predictive modelling activity is to set up a scenario that would accurately describe the outcomes of a given student assuming no new intervention. For instance, one might use a predictive model to determine when a given individual is likely to complete their academic degree. Applying this model to individual students will provide insight into when they might complete their degrees assuming no intervention strategy is employed. Thus, while it is important for a predictive model to generate accurate scenarios, these models are not generally deployed without an intervention or remediation strategy in mind.

Strong candidate problems for a successful predictive modelling approach are those in which there are quantifiable characteristics of the subject being modelled, a clear outcome of interest, the ability to intervene in situ, and a large set of data. Most importantly, there must be a recurring need, such as a class being ordered year after year, where the historical data on learners (the training set) is indicative of future learners (the testing set).

Conversely, several factors make predictive modelling more difficult or less appropriate. For example, both sparse and noisy data present challenges when trying to create accurate predictive models. Data sparsity, or missing data, can occur for a variety of reasons, such as students choosing not to provide optional information. Noisy data occurs when a measurement fails to capture the intended data accurately, such as determining a student's location from their IP address when some students are using virtual private networks (proxies used to circumvent region restrictions, a not uncommon practice in countries such as China). Finally, in some domains, inferences produced by predictive models may be at odds with ethical or equitable practice, such as using models of student at-risk predictions to limit the admissions of said students (exemplified in Stripling et al., 2016).

Data Collection

In predictive modelling, historical data is used to generate models of relationships between features. One of the first activities for a researcher is to identify the outcome variable (e.g., grade or achievement level) as well as the suspected correlates of this variable (e.g., gender, ethnicity, access to given resources). Given the situational nature of the modelling activity, it is important to choose only those correlates available at or before the time in which an intervention might be employed. For instance, a midterm examination grade might be predictive of a final grade in the course, but if the intent is to intervene before the midterm, this data value should be left out of the modelling activity.

In time-based modelling activities, such as the prediction of a student final grade, it is common for multiple models to be created (e.g., Barber & Sharkey, 2012), each corresponding to a different time period and set of observed variables. For instance, one might generate predictive models for each week of the course, incorporating into each model the results of weekly quizzes, student demographics, and the amount of engagement the students have had with respect digital resources to date in the course.

While state-based data, such as data about demographics (e.g., gender, ethnicity), relationships (e.g., course enrollments), psychological measures (e.g., grit, as in Duckworth, Peterson, Matthews, & Kelly, 2007, and aptitude tests), and performance (e.g., standardized test scores, grade point averages) are important for educational predictive models, it is the recent rise of big event-driven data collections that has been a particularly powerful enabler of predictive models (see Alhadad et al., 2015 for a deeper discussion). Event-data is largely student activity-based, and is derived from the learning technologies that students interact with, such as learning content management systems, discussion forums, active learning technologies, and video-based instructional tools. This data is large and complex (often in the order of millions of database rows for a single course), and requires significant effort to convert into meaningful features for machine learning.

Of pragmatic consideration to the educational researcher is obtaining access to event data and creating the necessary features required for the predictive modelling process. The issue of access is highly context-specific and depends on institutional policies and processes as well as governmental restrictions (such as FERPA in the United States). The issue of converting complex data (as is the case with event-based data) into features suitable for predictive modelling is referred to as *feature engineering*, and is a broad area of research itself.

Classification and Regression

In statistical modelling, there are generally four types of data considered: categorical, ordinal, interval, and ratio. Each type of data differs with respect to the kinds of relationships, and thus mathematical operations, which can be derived from individual elements. In practice, ordinal variables are often treated as categorical, and interval and ratio are considered as numeric. Categorical values may be binary (such as predicting whether a student will pass or fail a course) or multivalued (such as predicting which of a given set of possible practice questions would be most appropriate for a student). Two distinct classes of algorithms exist for these applications; classification algorithms are used to predict categorical values, while regression algorithms are used to predict numeric values.

Feature Selection

In order to build and apply a predictive model, features that correlate with the value to predict must be created. When choosing what data to collect, the practitioner should err on the side of collecting more information rather than less, as it may be difficult or impossible to add additional data later, but removing information is typically much easier. Ideally, there would be some single feature that perfectly correlates with the chosen outcome prediction. However, this rarely occurs in practice. Some learning algorithms make use of all available attributes to make predictions, whether they are highly informative or not, whereas others apply some form of variable selection to eliminate the uninformative attributes from the model.

Depending on the algorithm used to build a predictive model, it can be beneficial to examine the correlation between features, and either remove highly correlated attributes (the multicollinearity problem in regression analyses), or apply a transformation to the features to eliminate the correlation. Applying a learning algorithm that naively assumes independence of the attributes can result in predictions with an over-emphasis on the repeated or correlated features. For instance, if one is trying to predict the grade of a student in a class and uses an attribute of both attendance in-class on a given day as well as whether a student asked a question on a given day, it is important for the researcher to acknowledge that the two features are not independent (e.g., a student could not ask a question if they were not in attendance). In practice, the dependencies between features are often ignored, but it is important to note that some techniques used to clean and manipulate data may rely upon an assumption of independence.[8] By determining an informative subset of the features, one can reduce the computational complexity of the predictive model, reduce data storage and collection requirements, and aid in simplifying predictive models for explanation.

[8] The authors share an anecdote of an analysis that fell prey to the dangers of assuming independence of attributes when using resampling techniques to boost certain classes of data when applying the synthetic minority over-sampling technique (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). In that case, missing data with respect to city and province resulted in a dataset containing geographically impossible combinations, reducing the effectiveness of the attributes and lowering the accuracy of the model.

Missing values in a dataset may be dealt with in several ways, and the approach used depends on whether data is missing because it is unknown or because it is not applicable. The simplest approach either is to remove the attributes (columns) or instances (rows) that have missing values. There are drawbacks to both of these techniques. For example, in domains where the total amount of data is quite small, the impact of removing even a small portion of the dataset can be significant, especially if the removal of some data exacerbates an existing class imbalance. Likewise, if all attributes have a small handful of missing values, then attribute removal will remove all of the data, which would not be useful. Instead of deleting rows or columns with missing data, one can also infer the missing values from the other known data. One approach is to replace missing values with a "normal" value, such as the mean of the known values. A second approach is to fill in missing values in records by finding other similar records in the dataset, and copying the missing values from their records.

The impact of missing data is heavily tied to the choice of learning algorithm. Some algorithms, such as the naïve Bayes classifier can make predictions even when some attributes are unknown; the missing attributes are simply not used in making a prediction. The nearest neighbour classifier relies on computing the distance between two data points, and in some implementations the assumption is made that the distance between a known value and a missing value is the largest possible distance for that attribute. Finally, when the C4.5 decision tree algorithm encounters a test on an instance with a missing value, the instance is divided into fractional parts that are propagated down the tree and used for a weighted voting. In short, missing data is an important consideration that both regularly occurs and is handled differently depending upon the machine learning method and toolkit employed.

Methods for Building Predictive Models

After collecting a dataset and performing attribute selection, a predictive model can be built from historical data. In the most general terms, the purpose of a predictive model is to make a prediction of some unknown quantity or attribute, given some related known information. This section will briefly introduce several such methods for building predictive models. A fundamental assumption of predictive modelling is that the relationships that exist in the data gathered in the past will still exist in the future. However, this assumption may not hold up in practice. For example, it may be the case that (according to the historical data collected) a student's grade in Introductory Calculus is highly correlated with their likelihood of completing a degree within 4 years. However, if there is a change in the instructor of the course, the pedagogical technique employed, or the degree programs requiring the course, this course may no longer be as predictive of degree completion as was originally thought. The practitioner should always consider whether patterns discovered in historical data should be expected in future data.

A number of different algorithms exist for building predictive models. With educational data, it is common to see models built using methods such as these:

1. **Linear Regression** predicts a continuous numeric output from a linear combination of attributes.
2. **Logistic Regression** predicts the odds of two or more outcomes, allowing for categorical predictions.
3. **Nearest Neighbours Classifiers** use only the closest labelled data points in the training dataset to determine the appropriate predicted labels for new data.
4. **Decision Trees** (e.g., C4.5 algorithm) are repeated partitions of the data based on a series of single attribute "tests." Each test is chosen algorithmically to maximize the purity of the classifications in each partition.
5. **Naïve Bayes Classifiers** assume the statistical independence of each attribute given the classification, and provide probabilistic interpretations of classifications.
6. **Bayesian Networks** feature manually constructed graphical models and provide probabilistic interpretations of classifications.
7. **Support Vector Machines** use a high dimensional data projection in order to find a hyperplane of greatest separation between the various classes.
8. **Neural Networks** are biologically inspired algorithms that propagate data input through a series of sparsely interconnected layers of computational nodes (neurons) to produce an output. Increased interest has been shown in neural network approaches under the label of *deep learning*.
9. **Ensemble Methods** use a voting pool of either homogeneous or heterogeneous classifiers. Two prominent techniques are bootstrap aggregating, in which several predictive models are built from random sub-samples of the dataset, and boosting, in which successive predictive models are designed to account for the misclassifications of the prior models.

Most of these methods, and their underlying software implementations, have tunable parameters that change the way the algorithm works depending upon expectations of the dataset. For instance, when building decision trees, a researcher might set a minimum leaf size or maximum depth of tree parameter used in order to ensure some level of generalizability.

Numerous software packages are available for the building of predictive modelling, and choosing the right package depends highly on the researcher's experience, the desired classification or regression approach, and the amount of data and data cleaning required. While a comprehensive discussion of these platforms is outside the scope of this chapter, the freely available and open-source package Weka (Hall et al., 2009) provides implementations of a number of the previously mentioned modelling methods, does not require programming knowledge to use, and has associated educational materials including a textbook (Witten, Frank, & Hall, 2011) and series of free online courses (Witten, 2016).

While the breadth of techniques covered within a given software package has led to it being commonplace for researchers (including educational data scientists) to publish tables of classification accuracies for a number of different methods, the authors caution against this. Once a given technique has shown promise, time is better spent reflecting on the fundamental assumptions of classifiers (e.g., with respect to missing data or dataset imbalance), exploring ensembles of classifiers, or tuning the parameters of particular methods being employed. Unless the intent of the research activity is to compare two statistical modelling approaches specifically, educational data scientists are better off tying their findings to new or existing theoretical constructs, leading to a deepening of understanding of a given phenomenon. Sharing data and analysis scripts in an open science fashion provides better opportunity for small technique iterations than cluttering a publication with tables of (often) uninteresting precision and recall values.

Evaluating a Model

In order to assess the quality of a predictive model, a test dataset with known labels is required. The predictions made by the model on the test set can be compared to the known true labels of the test set in order to assess the model. A wide variety of measures is available to compare the similarity of the known true labels and the predicted labels. Some examples include prediction accuracy (the raw fraction of test instances correctly classified), precision, and recall.

Often, when approaching a predictive modelling problem, only one omnibus set of data is available for building. While it may be tempting to reuse this same dataset as a test set to assess model quality, the performance of the predictive model will be significantly higher on this dataset than would be seen on a novel dataset (referred to as overfitting the model). Instead, it is common practice to "hold out" some fraction of the dataset and use it solely as a test set to assess model quality.

The simplest approach is to remove half of the data and reserve it for testing. However, there are two drawbacks to this approach. First, by reserving half of the data for testing, the predictive model will only be able to make use of half of the data for model fitting. Generally, model accuracy increases as the amount of available data increases. Thus, training using only half of the available data may result in predictive models with poorer performance than if all the data had been used. Second, our assessment of model quality will only be based on predictions made for half of the available data. Generally, increasing the number of instances in the test set would increase the reliability of the results. Instead of simply dividing the data into training and testing partitions, it is common to use a process of k-fold cross validation in which the dataset is partitioned at random into k segments; k distinct predictive models are constructed, with each model training on all but one of the segments, and testing on the single held out segment. The test results are then pooled from all k test segments, and an assessment of model quality can be performed. The important benefits of k-fold cross validation are that every available data point can be used as part of the test set, no single data point is ever used in both the training set and test set of the same classifier at the same time, and the training sets used are nearly as large as all of the available data.

An important consideration when putting predictive modelling into practice is the similarity between the data used for training the model and the data available when predictions need to be made. Often in the educational domain, predictive models are constructed using data from one or more time periods (e.g., semesters or years), and then applied to student data from the next time period. If the features used to construct the predictive model include factors such as students' grades on individual assignments, then the accuracy of the model will depend on how similar the assignments are from one year to the next. To get an accurate assessment of model performance, it is important to assess the model in the same manner as will be used in situ. Build the predictive model using data available from one year, and then construct a testing set consisting of data from the following year, instead of dividing data from a single year into training and testing sets.

PREDICTIVE ANALYTICS IN PRACTICE

Predictive analytics are being used within the field of teaching and learning for many purposes, with one significant body of work aimed at identifying students at risk in their academic programming. For instance, Aguiar et al. (2015) describe the use of predictive models to determine whether students will graduate from secondary school on time, demonstrating how the accuracy of predictions changes as students advance from primary school through into secondary school. Predicted outcomes vary widely, and might include a specific summative grade or grade distribution for a student or class of achievement (Brooks et al., 2015) in a course. Baker, Gowda, and Corbett (2011) describe a method that predicts a formative achievement for a student based on their previous interactions with an intelligent tutoring system. In lower-risk and semi-formal settings such as massive open online courses (MOOCs), the chance that a learner might disengage from the learning activity mid-course is another heavily studied outcome (Xing, Chen, Stein, & Marcinkowski, 2016; Taylor, Veeramachaneni, & O'Reilly, 2014).

Beyond performance measures, predictive models have been used in teaching and learning to detect learners who are engaging in off-task behaviour (Xing and Goggins, 2015; Baker, 2007) such as "gaming the system" in order to answer questions correctly without learning (Baker, Corbett, Koedinger, & Wagner, 2004). Psychological constructs such as affective and emotional states have also been predictively modelled (D'Mello, Craig, Witherspoon, McDaniel, & Graesser, 2007; Wang, Heffernan, & Heffernan, 2015), using a variety of underlying data as features, such as textual discourse or facial characteristics. More examples of some of the ways predictive modelling has been used in Educational Data Mining in particular can be found in Koedinger, D'Mello, McLaughlin, Pardos, and Rosé (2015).

CHALLENGES AND OPPORTUNITIES

Computational and statistical methods for predictive modelling are mature, and over the last decade, a number of robust tools have been made available for educational researchers to apply predictive modelling to teaching and learning data. Yet a number of challenges and opportunities face the learning analytics community when building, validating, and applying predictive models. We identify three areas that could use investment in order to increase the impact that predictive modelling techniques can have:

1. *Supporting non-computer scientists in predictive modelling activities*. The learning analytics field is highly interdisciplinary and educational researchers, psychometricians, cognitive and social psychologists, and policy experts tend to have strong backgrounds in explanatory modelling. Providing support in the application of predictive modelling techniques, whether through the innovation of user-friendly tools or the development of educational resources on predictive modelling, could further diversify the set of educational researchers using these techniques.
2. *Creating community-led educational data science challenge initiatives*. It is not uncommon for researchers to address the same general theme of work but use slightly different datasets, implementations, and outcomes and, as such, have results that are difficult to compare. This is exemplified in recent predictive modelling research regarding dropout in massive open online courses, where a number of different authors (e.g., Brooks et al., 2015; Xing et al., 2016; Taylor et al., 2014; Whitehill, Williams, Lopez, Coleman, & Reich, 2015) have all done work with different datasets, outcome variables, and approaches.

   Moving towards a common and clear set of outcomes, open data, and shared implementations in order to compare the efficacy of techniques and the suitability of modelling methods for given problems could be beneficial for the community. This approach has been valuable in similar research fields and the broader data science community and we believe that educational data science challenges could help to disseminate predictive modelling knowledge throughout the educational research community while also providing an opportunity for the development of novel interdisciplinary methods, especially related to feature engineering.
3. Engaging in second order predictive modelling. In the context of learning analytics, we define second order predictive models as those that include historical knowledge as to the effects of and intervention in the model itself. Thus a predictive model that used student interactions with content to determine drop out (for instance) would be an example of first order predictive modelling, while a model that also includes historical data as to the effect of an intervention (such as an email prompt or nudge) would be considered a second order predictive model. Moving towards the modelling of intervention effectiveness is important when multiple interventions are available and personalized learning paths are desired.

Despite the multidisciplinary nature of the learning analytics and educational data mining communities, there is still a significant need for bridging understanding between the diverse scholars involved. An interesting thematic undercurrent at learning analytics conferences are the (sometimes-heated) discussions of the roles of theory and data as drivers of educational research. Have we reached the point of "the end of theory" (Anderson, 2008) in educational research? Unlikely, but this question is most salient within the subfield of predictive modelling in teaching and learning: while for some researchers the goal is understanding cognition and learning processes, others are interested in predicting future events and success as accurately as possible. With predictive models becoming increasingly complex and incomprehensible by an individual (essentially black boxes), it is important to start discussing more explicitly the goals of research agendas in the field, to better drive methodological choices between explanatory and predictive modelling techniques.

REFERENCES

Aguiar, E., Lakkaraju, H., Bhanpuri, N., Miller, D., Yuhas, B., & Addison, K. L. (2015). Who, when, and why: A machine learning approach to prioritizing students at risk of not graduating high school on time. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 93–102). New York: ACM.

Alhadad, S., Arnold, K., Baron, J., Bayer, I., Brooks, C., Little, R. R., Rocchio, R. A., Shehata, S., & Whitmer, J. (2015, October 7). The predictive learning analytics revolution: Leveraging learning data for student success. Technical report, EDUCAUSE Center for Analysis and Research.

Anderson, C. (2008, June 23). The end of theory: The data deluge makes the scientific method obsolete. Wired. https://www.wired.com/2008/06/pb-theory/

Baker. R. S. J. d. (2007). Modeling and understanding students' on-task behaviour in intelligent tutoring systems. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '07), 28 April–3 May 2007, San Jose, CA (pp. 1059–1068). New York: ACM.

Baker, R. S. J. d., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). On-task behaviour in the cognitive tutor classroom: When students game the system. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '04), 24–29 April 2004, Vienna, Austria (pp. 383–390). New York: ACM.

Baker, R. S. J. d., Gowda, S. M., & Corbett, A. T. (2011). Towards predicting future transfer of learning. *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (AIED '11), 28 June–2 July 2011, Auckland, New Zealand (pp. 23–30). Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Barber, R., & Sharkey, M. (2012). Course correction: Using analytics to predict course success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 259–262). New York: ACM. doi:10.1145/2330601.2330664

Brooks, C., Thompson, C., & Teasley, S. (2015). A time series interaction analysis method for building predictive models of learners using log data. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 126–135). New York: ACM.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.

D'Mello, S. K., Craig, S. D., Witherspoon, A., McDaniel, B., & Graesser, A. (2007). Automatic detection of learner's affect from conversational cues. *User Modeling and User-Adapted Interaction*, 18(1–2), 45–80.

Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The Weka data mining software: An update. *SIGKDD Explorations Newsletter*, 11(1), 10–18. doi:10.1145/1656274.1656278.

Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333–353.

Lonn, S., & Teasley, S. D. (2014). Student explorer: A tool for supporting academic advising at scale. *Proceedings of the 1st ACM Conference on Learning @ Scale* (L@S 2014), 4–5 March 2014, Atlanta, Georgia, USA (pp. 175–176). New York: ACM. doi:10.1145/2556325.2567867

Shmueli, G. (2010). To explain or to predict? *Statistical Science, 25*(3), 289–310. doi:10.1214/10-STS330

Stripling, J., Mangan, K., DeSantis, N., Fernandes, R., Brown, S., Kolowich, S., McGuire, P., & Hendershott, A. (2016, March 2). Uproar at Mount St. Mary's. The Chronicle of Higher Education. http://chronicle.com/specialreport/Uproar-at-Mount-St-Marys/30.

Taylor, C., Veeramachaneni, K., & O'Reilly, U.-M. (2014, August 14). Likely to stop? Predicting stopout in massive open online courses. http://dai.lids.mit.edu/pdf/1408.3382v1.pdf

Wang, Y., Heffernan, N. T., & Heffernan, C. (2015). Towards better affect detectors: Effect of missing skills, class features and common wrong answers. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 31–35). New York: ACM.

Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout. In O. C. Santos et al. (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. XXX–XXX). International Educational Data Mining Society. http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_112.pdf

Witten, I. H. (2016). Weka courses. The University of Waikato. https://weka.waikato.ac.nz/explorer

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*, 3rd ed. San Francisco, CA: Morgan Kaufmann Publishers.

Xing, W., Chen, X., Stein, J., & Marcinkowski, M. (2016). Temporal predication of dropouts in MOOCs: Reaching the low-hanging fruit through stacking generalization. *Computers in Human Behavior*, 58, 119–129.

Xing, W., & Goggins, S. (2015). Learning analytics in outer space: A hidden naive Bayes model for automatic students' on-task behaviour detection. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 176–183). New York: ACM.

Chapter 6: Going Beyond Better Data Prediction to Create Explanatory Models of Educational Data

Ran Liu, Kenneth R. Koedinger

School of Computer Science, Carnegie Mellon University, USA

DOI: 10.18608/hla17.006

ABSTRACT

In the statistical modelling of educational data, approaches vary depending on whether the goal is to build a predictive or an explanatory model. Predictive models aim to find a combination of features that best predict outcomes; they are typically assessed by their accuracy in predicting held-out data. Explanatory models seek to identify interpretable causal relationships between constructs that can be either observed or inferred from the data. The vast majority of educational data mining research has focused on achieving predictive accuracy, but we argue that the field could benefit from more focus on developing explanatory models. We review examples of educational data mining efforts that have produced explanatory models and led to improvements to learning outcomes and/or learning theory. We also summarize some of the common characteristics of explanatory models, such as having parameters that map to interpretable constructs, having fewer parameters overall, and involving human input early in the model development process.

Keywords: Explanatory models, model interpretability, educational data mining (EDM), closing the loop, cognitive models

Across the vast majority of educational data mining research, models are evaluated based on their predictive accuracy. Most often, this takes the form of assessing the model's ability to correctly predict successes and failures in a set of student response outcomes. Much less commonly, models may be validated based on their ability to predict post-test outcomes (e.g., Corbett & Anderson, 1995) or pre-test/post-test gains (e.g., Liu & Koedinger, 2015).

While predictive modelling has much to recommend it, the field of educational data mining could benefit from more emphasis on developing explanatory models. Explanatory models seek to identify interpretable constructs that are causally related to outcomes (Shmueli, 2010). In doing so, they provide an explanation of the data that can be connected to existing theory. The focus is on *why* a model fits the data well rather than only *that* it fits well. Often, explanatory models provide an interpretation of the data that has implications for theory, practice, or both. Here, we review educational data mining efforts that have produced explanatory models and, in turn, can lead to improvements to learning outcomes and/or learning theory.

Educational data mining research has largely focused on developing two types of models: the statistical model and the cognitive model. Statistical models drive the outer loop of intelligent tutoring systems (VanLehn, 2006) based on observable features of students' performance as they learn. Cognitive models are representations of the knowledge space (facts, concepts, skills, et cetera) underlying a particular educational domain. The majority of the research reviewed here concerns cognitive model refinement and discovery. We also briefly review other examples of explanatory models outside the realm of cognitive model discovery that educational data mining research has produced.

COGNITIVE MODEL DISCOVERY

Cognitive models map knowledge components (i.e., concepts, skills, and facts; Koedinger, Corbett, & Perfetti, 2012) to problem steps or tasks on which student performance can be observed. This mapping provides a way for statistical models to make inferences about students' underlying knowledge based on their observable performance on different problem steps. Thus, cognitive models are an important basis for the instructional design of automated tutors and are important for accurate assessment of learning and knowledge. Better cognitive models lead to better predictions of what a student knows, allowing adaptive learning to work more efficiently. Traditional ways of constructing cognitive models (Clark, Feldon, van Merriënboer, Yates, & Early, 2008) include structured interviews, think-aloud protocols, rational analysis, and labelling by domain experts. These methods, however, require human input and are often time consuming. They are also subjective, and previous research (Nathan, Koedinger, & Alibali, 2001; Koedinger & McLaughlin, 2010) has shown that expert-engineered cognitive models often ignore content distinctions that are important for novice learners. Here, we review three examples of efforts to discover and refine cognitive models based on data-driven techniques that alleviate expert bias while reducing the load on human input.

For statistical modelling purposes, the work described here uses a simplification of a cognitive model composed of hypothesized knowledge components. A knowledge component (KC) is a fact, concept, or skill required to succeed at a particular task or problem step. We refer to this specialized form of a cognitive model as a KC model or, alternatively, a Q-matrix (Barnes, 2005). The statistical model we used to evaluate the predictive fit of data-driven cognitive model discoveries is a logistic regression model called the additive factors model (AFM; Cen, Koedinger, & Junker, 2006), a generalization of item-response theory to accommodate learning effects.

Data-Driven Cognitive Model Improvement

Difficulty factors assessment (DFA; e.g., Koedinger & Nathan, 2004) moves beyond expert intuition by using a data-driven knowledge decomposition process to identify problematic elements of a defined task. In other words, when one task is much harder than a closely related task, the difference implies a knowledge demand of the harder task that is not present in the easier one. Stamper and Koedinger (2011) illustrated a method that uses DFA, along with freely accessible educational data and built-in visualization tools on DataShop[1] (Koedinger et al., 2010), to identify and validate cognitive model improvements. The method for cognitive model refinement iterates through the following steps: 1) inspect learning curve visualizations and fitted AFM coefficient estimates for a given KC model, 2) identify problematic KCs and hypothesize changes to the KC model, 3) re-fit the AFM with the revised KC model and investigate whether the new model fits the data better.

Through manual inspection of the visualizations of a geometry dataset (Koedinger, Dataset 76 in DataShop[2]), potential improvements to the best existing KC model at the time were identified (Stamper & Koedinger, 2011). Most of the KCs in this model exhibited relatively smooth learning curves with a consistent decline in error rate. One KC in the original model, *compose-by-addition*, exhibited a particularly noisy curve with large spikes in error rate at certain opportunity counts. In addition, the AFM parameter estimates for the *compose-by-addition* KC suggested no apparent learning (the slope parameter estimate was very close to zero, and not because the performance was at ceiling). A bumpy learning curve and low slope estimate are indications of a poorly defined KC. One common cause for a poorly defined KC is that some of its constituent items require some knowledge demand that other items do not. In other words, the original KC should really be split into two different KCs. To improve the KC model, all *compose-by-addition* problem steps were examined, and domain expertise was applied to hypothesize about additional knowledge that might be required on certain steps. As a result, the *compose-by-addition* KC was split into three distinct KCs, and each of the 20 steps previously labelled with the *compose-by-addition* KC were relabelled accordingly. The revised model resulted in smoother learning curves and, when fit with the AFM, yielded significantly better predictions of student performance than the original KC model did. Although this KC model improvement was aided by visualizations resulting from fitting a statistical model, the actual improvements were generated manually and thus were readily interpretable.

The discovered KC model improvements had clear implications for revising instruction. Koedinger, Stamper, McLaughlin, and Nixon (2013) used the data-driven KC model improvements to generate a revised version of the Geometry Area tutor unit. Revisions included adding the newly discovered skills to the KC model driving adaptive learning, resulting in changes to knowledge tracing, and the creation of new tasks to target the new skills. In an A/B experiment, half of the students completed the revised tutor unit and the other half

[1] http://pslcdatashop.org
[2] Geometry Area 1996–1997: https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=76

competed the original tutor unit. Students using the revised tutor reached mastery more efficiently and exhibited better learning on the skills targeted by the KC model improvement, based on pre- to post-test gains (Koedinger et al., 2013). These results show that the data-driven DFA technique lends itself to generating explanatory KC model refinements that can result in instructional modifications and improved learning outcomes.

Learning Factors Analysis

Learning factors analysis (LFA; Cen et al., 2006) was developed to automate the data-driven method of KC model refinement to further alleviate demands on human time. LFA searches across hypothesized knowledge components drawn from different existing KC models, evaluates different models based on their fit to data, and outputs the best-fitting KC model in the form of a symbolic model. As such, LFA greatly reduces demands on human effort while simultaneously easing the burden of interpretation, even if it does not automatically accomplish it.

We applied the LFA search process across 11 datasets spanning different domains and different educational technologies, all publicly available from DataShop. Across all 11 datasets, this automated discovery process improved KC models' fit to data beyond the best existing human-tagged KC models (Koedinger, McLaughlin, & Stamper, 2012). Importantly, we demonstrated in an example dataset (Koedinger, Dataset 76 in DataShop) an interpretable explanation for the specific improvements made by the best LFA-discovered model. A manual KC model comparison between the best-fitting LFA model and the best-fitting human-tagged model revealed that the LFA model tagged separate KCs for forwards (i.e., find area given radius) and backwards (i.e., find radius given area) circle area problems, whereas these had been grouped together as a single "circle-area" KC in the human-tagged model. No such differences were found between the models for other shapes like rectangles, triangles, and parallelograms. Applying domain expertise to interpret the automated discovery, we hypothesized that LFA's model improvement may have captured the difficulty of knowing when and how to apply a square root operation for backwards circle-area problems, which is not required for forwards circle-area problems nor for the backwards area problems of other shapes.

We then assessed the external validity of this interpretation beyond the dataset from which the discoveries were made. We evaluated the presence of the square root difficulty in a novel dataset (Bernacki, Dataset 748 in DataShop[3]), one with a different structure from that used to make the discovery (Liu, Koedinger, & McLaughlin, 2014). Among other differences, the novel dataset contained more backwards circle-area problems and, importantly, forwards (i.e., find area given side length) and backwards (i.e., find side length given area) *square-area* problems. These square-area problems were not at all present in the original dataset from which the LFA-generated discovery was made. Applying our interpretation of the discovery, we constructed a KC model that tags separate forwards and backwards KCs only for shapes where backwards steps require computing a square root (squares, circles) but not for shapes where backwards steps don't (triangles, rectangles, parallelograms). When used in conjunction with the AFM, this KC model yielded the best fit to the novel dataset compared to several expert-tagged control KC models.

Since the novel dataset had a different structure from the original dataset, including differences relevant to the KC model discovery (i.e., existence of backwards square-area problems), it would not have been viable to apply directly the LFA-discovered KC model on this new dataset. Interpretation is necessary in order to test the generalizability of discoveries across contexts with non-identical structures. Furthermore, interpretations help anchor all subsequent data exploration and analyses to something meaningful that can then be translated into concrete improvements to instructional design. Our current research is "closing the loop" on this LFA-generated discovery by assessing learning outcomes resulting from a tutor redesigned around the improved KC model (Liu & Koedinger, submitted).

Automated Cognitive Model Discovery Using SimStudent

An alternative automated approach uses a state-of-the-art machine-learning agent, SimStudent, to discover cognitive models automatically without requiring existing ones. SimStudent is an intelligent agent that inductively learns knowledge, in the form of rules, by observing a tutor solve sample problems and by solving problems on its own and receiving feedback (Li, Matsuda, Cohen, & Koedinger, 2015). One of the benefits of SimStudent is that it can simulate features of novices' learning trajectories of which domain experts may not even be aware. Real students entering a course do not usually have substantial domain-specific prior knowledge, so a realistic model of human learning ought not to assume this knowledge is given. In addition, SimStudent can be used to test alternative models of human learning to see which best predicts human behaviour (MacLellan, Harpstead, Patel, & Koedinger, 2016). For several datasets spanning various domains, SimStudent generated cognitive models that fit the data better than the best human-generated cognitive

[3] Motivation for learning HS geometry 2012 (geo-pa): https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=748

models (Li et al., 2011; MacLellan et al., 2016).

The output of the SimStudent's learning takes the form of production rules (Newell & Simon, 1972), and each production rule essentially corresponds to one knowledge component (KC) in a KC model. Using data from an Algebra dataset (Booth & Ritter, Dataset 293 in DataShop[4]) and in conjunction with the AFM, Li and colleagues (2011) compared a KC model generated by SimStudent to a KC model generated by hand-coding actual students' actions within the tutor. The SimStudent-generated model better fit the actual student performance data than the human-generated model did.

More importantly, inspecting the differences between the SimStudent model and the human-generated model revealed interpretable features that explained the advantages of the SimStudent model. One example of such a difference is that SimStudent created distinct production rules (KCs) for division-based algebra problems of the form Ax=B, where both A and B are signed numbers, and for the form −x=A, where only A is a signed number. To solve Ax=B, SimStudent learns to simply divide both sides by the signed number A. But, since −x does not represent its coefficient (−1) explicitly, SimStudent must first recognize that −x translates to −1x, and then it can divide both sides by −1. The human-generated model predicts that both forms of division problems should have the same error rates. In fact, real students have greater difficulty making the correct move on steps like −x = 6 than on steps like −3x = 6. Within the same Algebra dataset, problems of the form Ax=B (average error rate = 0.28) are easier than problems of the form −x=A (average error rate = 0.72). SimStudent's split of division problems into two distinct KCs suggests that students should be tutored on two subsets of problems, one subset corresponding to the form Ax=B and one subset specifically for the form −x=A. Explicit instruction that highlights for students that −x is the same as −1x may be beneficial (Li et al., 2011).

We hypothesized that the interpretation of this particular SimStudent KC model discovery would generalize to novel problem types, just as the LFA-generated model discovery did. In a novel equation-solving dataset (Ritter, Dataset 317 in DataShop[5]), we tested whether the explicit vs. implicit coefficient distinction similarly applied to *combine like terms* problems. We looked at differences in performance for items of the form Ax + Bx = C, where both A, B, and C are signed numbers (explicit-coefficient items), and items where either A or B were equal to 1 or −1 with the coefficient perceptually absent (implicit-coefficient items). This analysis confirmed that explicit-coefficient items (average error rate = 0.35) are easier than implicit-coefficient items (average error rate = 0.45) among *combine like terms* problems. This new dataset not only replicated the original finding that SimStudent made on divide problems, but it also revealed that the finding generalizes to a separate procedural skill, *combine like terms*.

Fitting a KC model with separate KCs for the explicit- vs. implicit-coefficient forms of *combine like terms* items revealed a large improvement in predictive fit relative to a KC model with a single *combine like terms* KC. Furthermore, although the learning curves for both the explicit-coefficient divide and combine like terms KCs reflected smooth and decreasing error rates, the respective learning curves for implicit-coefficient *divide* and *combine like terms* items were both flat, with slopes close to zero. This suggests that students would benefit greatly from more practice on, and more explicit attention to problem steps involving implicit coefficients. Here, again, the explanatory power of the SimStudent KC model discovery made it possible to generalize the explanation to distinct problem types on which SimStudent was never trained.

Comparison to Other Work

Both LFA and SimStudent are capable of producing cognitive model discoveries that not only significantly improve predictive accuracy but are readily interpretable and, thus, explanatory. We have demonstrated that the interpretations yielded by these cognitive model discoveries generalize to novel problem types not present in the data from which the discoveries were made. Finally, they produce clear recommendations for revising instruction, even in contexts that are very different from those in which the original data were collected. These are all hallmarks of explanatory modelling efforts that move beyond simply improving predictive accuracy to have meaningful impact on learning theory and instruction.

The fact that methods like LFA are "human-in-the-loop" – that is, requiring input from a domain expert – has been cited as a limitation. In the case of LFA, one or more expert-tagged cognitive models are required initially in order to produce new model discoveries. We argue, however, that this "human-in-the-loop" feature leads the results of such modelling efforts to be explanatory. There have been a number of recent efforts to fully automate the process of discovering and/or improving cognitive models (González-Brenes & Mostow, 2012; Lindsey, Khajah, & Mozer, 2014). These methods have much to recommend, as they dramatically reduce demands on human time and produce competitive results in predictive accuracy. However, the resulting cognitive models of these efforts have

[4] Improving skill at solving equations via better encoding of algebraic concepts (2006–2008): https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=293

[5] Algebra I 2007–2008 (Equation Solving Units): https://pslcdatashop.web.cmu.edu/DatasetInfo?datasetId=317

not been interpreted or acted upon with respect to improving instruction.

Other modelling efforts, including a "human-in-the-loop" component like Ordinal SPARFA-Tag (Lan, Studer, Waters, & Baraniuk, 2013), have yielded considerably more interpretable cognitive models than many alternative methods. Although humans must do any final interpretation of modelling efforts, methods like LFA and Ordinal SPARFA-Tag greatly improve the likelihood of generating sensible resulting models by incorporating the human effort up front. In fact, comparing the original SPARFA model (Lan, Studer, Waters, & Baraniuk, 2014), which only incorporates concept tags post-hoc, to Ordinal SPARFA-Tag, which incorporates domain-expert concept tags in the model development process up front, shows that the latter model results in much more interpretable cognitive models.

More attention and effort towards generating interpretable cognitive models is, in our view, progress in the right direction. Nevertheless, as we have argued, expert labelling is still subject to biases and does not offer much to advance learning theory using the rich educational data available. Human involvement improves interpretability, whereas the data-driven component offers ways to alleviate subjective biases and advance our understanding of how novices learn. Methods such as LFA leverage both the unique strengths of human involvement and of automation towards creating models that are more predictive *and* explanatory.

STUDENT GROUPING

A growing body of research suggests that modelling student-specific variability in statistical models of educational data can yield better predictive accuracies and potentially inform instruction. Prior attempts to group students based on features available in educational datasets have focused on techniques such as K-means and spectral clustering. These techniques have been used to generate student clusters predictive of post-test performance (Trivedi, Pardos, & Heffernan, 2011) and that yield predictive accuracy improvements when clusters are fit with different sets of parameters (Pardos, Trivedi, Heffernan, & Sárközy, 2012). Many clustering techniques, however, tend to result in student groupings that are difficult to interpret. Yet, interpretation is critical if the results of clustering are to eventually inform improvements in instructional policy (e.g., individualizing instruction appropriately to different groups of students).

In recent research (Liu & Koedinger, 2015), we developed a method for grouping students that not only dramatically improves the predictive accuracy of the AFM but inherently lends itself to producing meaningful student groups. By doing a first-pass fit of the AFM to the data and examining systematic patterns in the residuals (differences between predicted and actual data) across different practice opportunities, we consistently found students belonging to one of three learning rate groups: 1) those who exhibit flatter learning curves than the AFM predicts, 2) those who exhibit steeper learning curves, and 3) those whose learning curves are on par with the model's predictions. Introducing a parameter that individualizes learning rates to each of these learning rate groups substantially improves model predictive accuracy, beyond that of the regular AFM, across a variety of datasets spanning multiple educational domains. Across datasets, the slope parameter estimates for each of the three groups were consistent with our interpretation of the groups (i.e., the estimated group-level slopes were always lowest for the flat-curve group, and highest for the steep-curve group). Furthermore, in a subset of datasets for which there exist paper pre- and post-test data, we observed a systematic relationship between learning-curve group and the degree of pre- to post-test improvement (Liu & Koedinger, 2015).

Unlike other, more "bottom-up" methods of creating stereotyped groups of students, this method yielded student groups that are readily interpretable and potentially actionable. For example, it is clear that the flat-curve student group represents either students who are already performing at ceiling when they start the unit or curriculum (and thus do not have much room for improvement) or students who are starting anywhere below ceiling but struggling to progress with the material. In either case, there are clear instructional implications for students classified into this group. The explanatory power of the resulting model again benefitted from doing some up-front interpretation and developing the model with an eye towards interpretability.

TOWARDS BUILDING EXPLANATORY MODELS

We argue for the importance of considering the interpretability and actionability of educational data mining efforts in producing more explanatory models. For a model to be explanatory, one should be able to understand *why* the model achieves better predictive accuracy than alternatives. In addition, the understanding of this *why* should either advance our understanding of how learners learn the relevant material or have clear implications for instructional improvements, or both. We summarize by outlining some of the features that tend to characterize explanatory models.

Explanatory modelling efforts tend to start with "clean" independent variables that have either simple functions or map to clearly defined constructs. For example, LFA derives new variables from existing, expert-labelled variables using simple split, merge, or add operators. Another example comes from automated analyses of verbal data in education, a branch of educational data mining that includes automated essay scoring, producing tutorial dialogue, and computer-supported collaborative learning. A major consideration in this area is how to transform raw text or transcriptions into features that can be used in a machine-learning algorithm. Approaches to this issue range from simple "bag of words" methods, which counts the frequency of each word present in the text, to much more sophisticated linguistic analyses. One consistent theme across findings is that feature representations motivated by interpretable, theoretical frameworks have been among the most promising (Rosé & Tovares, in press; Rosé & VanLehn, 2005). Thus, incorporating some human time and thought into defining and labelling these independent variables up front can greatly improve the explanatory power of the resulting model.

Another feature of explanatory models, one that relates most to actionability, is that the dependent variable maps to a well-defined construct. The work on learning rate groups is an example of this: since the groups to which students are classified are defined up front, it is clear what it means for a student to be in the "flat" learning curve group, as opposed to the "steep" one. This makes the results from modelling readily actionable. Another body of research in which the dependent variable tends to be well mapped to an interpretable construct is the modelling of affect and motivation using features of tutor log data. These techniques use pre-defined psychological or behavioural constructs, measured through questionnaires or expert observations, to develop and refine "detectors" that can identify those constructs within tutor log data activity (e.g., Winne & Baker, 2013; San Pedro, Baker, Bowers, & Heffernan, 2013; D'Mello, Blanchard, Baker, Ocumpaugh, & Brawner, 2014). The "detectors" are developed specifically to identify pre-determined constructs and, thus, the results of these algorithms are readily actionable. For example, Affective AutoTutor is an intelligent tutoring system for computer literacy that automatically models students' confusion, frustration, and boredom in real time. Detection of these affective states is then used to adapt the tutor actions in a manner that responds accordingly. An experimental study "closing the loop" on this affective detector showed higher learning gains for low-domain knowledge students who interacted with the Affective AutoTutor compared to a non-affective version (D'Mello et al., 2010). For these modelling efforts to be fully explanatory though, interpretations of the independent variables driving the affective outcomes are also needed.

Finally, explanatory models tend to be characterized by fewer estimated parameters (independent variables, or features). For example, the AFM has only one parameter for each student and two parameters for each knowledge component. Adding learning rate groups extends the model by only one additional parameter, group membership. This makes the contribution of the added parameter easy to attribute and interpret. Having fewer parameters also allows each parameter's estimates to have more explanatory power, alleviating issues of indeterminacy. Because the AFM has only one difficulty parameter and one learning parameter for each KC, one can, for example, meaningfully interpret a low learning parameter estimate as suggesting that KC needs either refinement or instructional improvement.

We have illustrated some ways in which concrete steps in the design of educational data modelling efforts can yield more explanatory models. The relationships between the fields of educational data mining, learning theory, and the practice of education could be greatly strengthened with increased attention to the explanatory power of models and their ability to influence future learning outcomes.

REFERENCES

Barnes, T. (2005). The Q-matrix method: Mining student response data for knowledge. *Proceedings of* AAAI 2005: *Educational Data Mining Workshop* (pp. 39–46). Technical Report WS-05-02. Menlo Park, CA: AAAI Press. http://www.aaai.org/Library/Workshops/ws05-02.php

Cen, H., Koedinger, K. R., & Junker, B. (2006). Learning factors analysis: A general method for cognitive model evaluation and improvement. In M. Ikeda, K. Ashlay, T.-W. Chan (Eds.), *Proceedings of the 8th International Conference on Intelligent Tutoring Systems* (ITS 2006), 26–30 June 2006, Jhongli, Taiwan (pp. 164–175). Berlin: Springer-Verlag.

Clark, R. E., Feldon, D., van Merriënboer, J., Yates, K., & Early, S. (2008). Cognitive task analysis. In J. M. Spector, M. D. Merrill, J. van Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling & User-Adapted Interaction*, 4, 253–278.

D'Mello, S., Blanchard, N., Baker, R., Ocumpaugh, J., & Brawner, K. (2014). I feel your pain: A selective review of affect sensitive instructional strategies. In R. Sottilare, A. Graesser, X. Hu, & B. Goldberg (Eds.), *Design recommendations for adaptive intelligent tutoring systems: Adaptive instructional strategies* (Vol. 2). Orlando, FL: US Army Research Laboratory.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., Perkins, L., & Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* (ITS 2010), 14–18 June 2010, Pittsburgh, PA, USA (pp. 245–254). Springer.

González-Brenes, J. P., & Mostow, J. (2012). Dynamic cognitive tracing: Towards unified discovery of student and cognitive models. In K. Yacef, O. Zaïane, A. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (EDM2012), 19–21 June, 2012, Chania, Greece (pp. 49–56). International Educational Data Mining Society.

Koedinger, K. R., Baker, R. S. J. d., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J. (2010). A data repository for the EDM community: The PSLC DataShop. In C. Romero, S. Ventura, M. Pechenizkiy, & R. S. J. d. Baker (Eds.), *Handbook of educational data mining*. Boca Raton, FL: CRC Press.

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36(5), 757–798.

Koedinger, K. R., & McLaughlin, E. A. (2010). Seeing language learning inside the math: Cognitive analysis yields transfer. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (CogSci 2010), 11–14 August 2010, Portland, OR, USA (pp. 471–476). Austin, TX: Cognitive Science Society.

Koedinger, K. R., McLaughlin, E. A., & Stamper, J. C. (2012). Automated cognitive model improvement. In K. Yacef, O. Zaïane, A. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (EDM2012), 19–21 June, 2012, Chania, Greece (pp. 17–24). International Educational Data Mining Society.

Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*, 13(2), 129–164.

Koedinger, K. R., Stamper, J. C., McLaughlin, E. A., & Nixon, T. (2013). Using data-driven discovery of better cognitive models to improve student learning. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Proceedings of the 16th International Conference on Artificial Intelligence in Education* (AIED '13), 9–13 July 2013, Memphis, TN, USA (pp. 421–430). Springer.

Lan, A. S., Studer, C., Waters, A. E., & Baraniuk, R. G. (2013). Tag-aware ordinal sparse factor analysis for learning and content analytics. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July, Memphis, TN, USA (pp. 90–97). International Educational Data Mining Society/Springer.

Lan, A. S., Studer, C., Waters, A. E., & Baraniuk, R. G. (2014). Sparse factor analysis for learning and content analytics. *Journal of Machine Learning Research*, 15, 1959–2008.

Li, N., Cohen, W., Koedinger, K. R., & Matsuda, N. (2011). A machine learning approach for automatic student model discovery. In M. Pechenizkiy et al. (Eds.), *Proceedings of the 4th International Conference on Education Data Mining* (EDM2011), 6–8 July 11, Eindhoven, Netherlands (pp. 31–40). International Educational Data Mining Society.

Li, N., Matsuda, N., Cohen, W. W., & Koedinger, K. R. (2015). Integrating representation learning and skill learning in a human-like intelligent agent. *Artificial Intelligence*, 219, 67–91.

Lindsey, R. V., Khajah, M., & Mozer, M. C. (2014). Automatic discovery of cognitive skills to improve the prediction of student learning. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberge (Eds.), *Advances in Neural Information Processing Systems*, 27, 1386–1394. La Jolla, CA: Curran Associates Inc.

Liu, R., & Koedinger, K. R. (submitted). Closing the loop: Automated data-driven skill model discoveries lead to improved instruction and learning gains. *Journal of Educational Data Mining*.

Liu, R., & Koedinger, K. R. (2015). Variations in learning rate: Student classification based on systematic residual error patterns across practice opportunities. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Education Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 420–423). International Educational Data Mining Society.

Liu, R., Koedinger, K. R., & McLaughlin, E. A. (2014). Interpreting model discovery and testing generalization to a new dataset. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 107–113). International Educational Data Mining Society.

MacLellan, C. J., Harpstead, E., Patel, R., & Koedinger, K. R. (2016). The apprentice learner architecture: Closing the loop between learning theory and educational data. In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (EDM2016), 29 June–2 July 2016, Raleigh, NC, USA (pp. 151–158). International Educational Data Mining Society.

Nathan, M. J., Koedinger, K. R., & Alibali, M. W. (2001). Expert blind spot: When content knowledge eclipses pedagogical content knowledge. In L. Chen et al. (Eds.), *Proceedings of the 3rd International Conference on Cognitive Science* (pp. 644–648). Beijing, China: USTC Press. http://pact.cs.cmu.edu/pubs/2001_NathanEtAl_ICCS_EBS.pdf

Newell, A., & Simon, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.

Pardos, Z. A., Trivedi, S., Heffernan, N. T., & Sárközy, G. N. (2012). Clustered knowledge tracing. In S. A. Cerri, W. J. Clancey, G. Papadourakis, K.-K. Panourgia (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (ITS 2012), 14–18 June 2012, Chania, Greece (pp. 405–410). Springer.

Rosé, C. P., & Tovares, A. (in press). What sociolinguistics and machine learning have to say to one another about interaction analysis. In L. Resnick, C. Asterhan, & S. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

Rosé, C. P, & VanLehn, K. (2005). An evaluation of a hybrid language understanding approach for robust selection of tutoring goals. *International Journal of Artificial Intelligence in Education*, 15, 325–355.

San Pedro, M., Baker, R. S., Bowers, A. J., & Heffernan, N. T. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July, Memphis, TN, USA (pp. 177–184). International Educational Data Mining Society/Springer.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*, 25(3), 289–310. doi:10.1214/10-STS330

Stamper, J., & Koedinger, K. R. (2011). Human-machine student model discovery and improvement using data. *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (AIED '11), 28 June–2 July, Auckland, New Zealand (pp. 353–360). Springer.

Trivedi, S., Pardos, Z. A., & Heffernan, N. T. (2011). Clustering students to generate an ensemble to improve standard test score predictions. *Proceedings of the 15th International Conference on Artificial Intelligence in Education* (AIED '11), 28 June–2 July, Auckland, New Zealand (pp. 377–384). Springer.

VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16, 227–265.

Winne, P., & Baker, R. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, 5, 1–8.

Chapter 7: Content Analytics: The Definition, Scope, and an Overview of Published Research

Vitomir Kovanović[1], Srećko Joksimović[2], Dragan Gašević[1,2], Marek Hatala[3], George Siemens[4]

[1] School of Informatics, The University of Edinburgh, United Kingdom
[2] Moray House School of Education, The University of Edinburgh, United Kingdom
[3] School of Interactive Arts and Technology, Simon Fraser University, Canada
[4] LINK Research Lab, The University of Texas at Arlington, USA
DOI: 10.18608/hla17.007

ABSTRACT

The field of learning analytics recently attracted attention from educational practitioners and researchers interested in the use of large amounts of learning data for understanding learning processes and improving learning and teaching practices. In this chapter, we introduce *content analytics* – a particular form of learning analytics focused on the analysis of different forms of educational content. We provide the definition and scope of content analytics and a comprehensive summary of the significant content analytics studies in the published literature to date. Given the early stage of the learning analytics field, the focus of this chapter is on the key problems and challenges for which existing content analytics approaches are suitable and have been successfully used in the past. We also reflect on the current trends in content analytics and their position within a broader domain of educational research.

Keywords: Content analytics, learning content

With the large amounts of data related to student learning being collected by digital systems, the potential for using this data for improving learning processes and teaching practices is widely recognized (Gašević, Dawson, & Siemens, 2015). The emerging field of learning analytics recently gained significant attention from educational researchers, practitioners, administrators, and others interested in the intersection of technology and education and the use of this vast amount of data for improving learning and teaching (Buckingham Shum & Ferguson, 2012). Among the different types of data, the analysis of learning content is commonly used for the development of learning analytics systems (Buckingham Shum & Ferguson, 2012; Chatti, Dyckhoff, Schroeder, & Thüs, 2012; Ferguson, 2012; Ferguson & Buckingham Shum, 2012). These include various forms of data produced by instructors (course syllabi, documents, lecture recordings), publishers (textbooks), or students (essays, discussion messages, social media postings). In this chapter, we introduce *content analytics*, an umbrella term used to refer to different types of learning analytics focusing on the analysis of various forms of learning content. We further provide a critical reflection on the state of the content analytics domain, identifying potential shortcomings and directions for future studies. We begin by discussing different forms of learning content and commonly adopted definitions of content analytics. Special attention is given to the range of problems commonly addressed by content analytics, as well as to various methodological approaches, tools, and techniques.

Learning Content and Content Analytics

According to Moore (1989), the defining characteristic of any form of education is the interaction between learners and learning content. Without content "there cannot be education since it is the process of intellectually interacting with the content that results in changes in the learner's understanding, the learner's perspective, or the cognitive structures of the learner's mind" (p. 2). While the most commonly used forms of educational content are written materials (Cook, Garside, Levinson, Dupras, & Montori, 2010), the ubiquitous access to personal computers and the Internet resulted in both a broad availability of learning resources and increased use of interactive and multimedia educational resources. Likewise, the emergence of web-based systems such as blogs and online discussion forums, and popular social media platforms (Twitter, Facebook) introduced a new dimension and provided access to a relatively new set of learner-generated resources (De Freitas, 2007, p. 16). The overall result is that landscape of educational content is expanding and diversifying, bringing along a new set of potential advantages, benefits, challenges, and risks (De Freitas, 2007). This global trend also creates fertile ground for the development of novel learning analytics approaches.

To provide an overview of content analytics literature, we should first define what is meant by content analytics. We define content analytics as

> Automated methods for examining, evaluating, indexing, filtering, recommending, and visualizing different forms of digital learning content, regardless of its producer (e.g., instructor, student) with the goal of understanding learning activities and improving educational practice and research.

This definition reveals that content analytics focuses on the automated analysis of the different "resources" (textbooks, web resources) and "products" (assignments, discussion messages) of learning. This is in clear contrast to analytics focused on the analysis of student behavioural data, such as the analysis of trace data from learning management systems. Although in general students can produce learning content of different types (text, video, audio), given the present state of educational technologies, and online/blended learning pedagogies, the content produced by the learners is predominantly text-based (assignment responses, discussion messages, essays). While there are cases where students produce non-textual content (video recordings of their presentations), they still represent a relative minority; consequently, very few analytical systems have been developed. Thus, the focus of this chapter is predominantly on text-based learning content, despite the broader definition of content analytics, which also encompasses multimedia learning content.

We should point out that content analytics is primarily defined in terms of the application domain, as many of the tools and techniques used are also employed in other types of learning analytics. As such, content analytics encompasses several more specific forms of analytics, including discourse analytics (Knight & Littleton, 2015), writing analytics (Buckingham Shum et al., 2016), assessment analytics (Ellis, 2013), and social learning analytics (Buckingham Shum & Ferguson, 2012). These particular analytics define their foci more specifically to examine learning content produced in particular learning products, processes, or contexts. As a consequence, our definition is broader than, for example, the definition of social content analytics by Buckingham Shum and Ferguson (2012), as a "variety of automated methods that can be used to examine, index and filter online media assets, with the intention of guiding learners through the ocean of potential resources available to them" (p. 15). We argue that the definition of content analytics used in this report – which does not focus on a particular learning setting or process – enables the development of standard analytical approaches applicable to many similar learning domains. Given the early stage of learning analytics development, the focus on the type of learning materials and the methodologies, techniques, and tools for their analysis promotes the establishment of community-wide standards of conducting content analytics research, which is critical for the advancement of the learning analytics field.

It is important to emphasize the difference between *content analysis* (Krippendorff, 2003) and content analytics, which are both commonly used techniques in educational research (Ferguson & Buckingham Shum, 2012). Despite similar names, content analysis is a much older and well-established research technique widely used across social sciences, including research in education, educational technology, and distance/online education (De Wever, Schellens, Valcke, & Van Keer, 2006; Donnelly & Gardner, 2011; Strijbos, Martens, Prins, & Jochems, 2006) to assess latent variables of written text. Given that many of the learning analytics systems are also focused on the examination of latent constructs, a large part of content analytics is an application of computational techniques for the purpose of content analysis (Kovanović, Joksimović, Gašević, & Hatala, 2014). However, content analytics includes different additional forms of analysis, which are not the focus of content analysis, such as assessment of student writings, automated student grading, or topic discovery in the document corpora.

CONTENT ANALYTICS TASKS AND TECHNIQUES

To provide an overview of content analytics, we conducted a review of the published literature on learning analytics and educational technology to identify research studies that made use of content analytics. We looked at the proceedings of the Learning Analytics and Knowledge Conference, the *Journal of Learning*

Analytics, the *Journal of Educational Data Mining*, the *Journal of Artificial Intelligence in Education*, and Google Scholar. After obtaining the relevant studies, we grouped them based on the research problems being addressed. We identified three groups of studies roughly focused on the three main types of data used for content analytics (i.e., learning resources, students' learning products, and students' social interactions). The remainder of this section provides a detailed overview of the identified groups of studies and associated tools and techniques.

Content Analytics of Learning Resources

One of the earliest uses of content analytics was for the analysis of educational resources and materials, and recommendation, organization, and evaluation of those resources. Given the vast amounts of learning materials available to students, one domain of particular interest is the recommendation of relevant learning-related content, based on various criteria such as student interest or course progress (Manouselis, Drachsler, Vuorikari, Hummel, & Koper, 2011; Romero & Ventura, 2010). The development of content analytics systems is typically based on recommender systems technologies, which can be split into two broad categories (Drachsler, Hummel, & Koper, 2008):

1. **Collaborative filtering** (CF) techniques, in which resources being recommended to a student were found by looking for either 1) *related students* (i.e., user-based CF), or 2) *related resources* (i.e., item-based CF). In the former case, students with a substantial overlap in their use of resources probably share common interests; in the latter case, resources used together by a large number of users are likely to be similar.
2. **Content-based** techniques, in which recommendations are discovered by directly comparing the content of resources to be recommended and by looking for most similar resources to the ones a student is currently using or that match the student's profile data.

Both approaches have been extensively used in educational technology (for an overview see Drachsler et al. 2008; Manouselis et al., 2011). For example, Walker, Recker, Lawless, and Wiley (2004) built AlteredVista, a collaborative system for discovering useful educational resources, while Zaldivar, García, Burgos, Kloos, and Pardo (2011) used content-based techniques to recommend course notes to students, based on their document browsing patterns. Content-based methods have also been used to recommend solutions (Hosseini & Brusilovsky, 2014) and relevant examples (Muldner & Conati, 2010) to programming tasks, and even to recommend suitable academic courses (Bramucci & Gaston, 2012). It should also be noted that the quality of recommendations is often dependent on the selection of particular document similarity measures (Verbert et al., 2012), which must be chosen to match the given learning context or activity.

Another important domain represents the automatic organization and classification of different instructional materials (often different learning objects), using automated techniques for keyword extraction, tagging, and clustering. For example, Bosnić, Verbert, and Duval (2010) compared different techniques for keyword extraction from learning objects, while Cardinaels, Meire, and Duval (2005) showed that an analysis of document content, usage, and context could be used to automatically create relevant metadata information for a given learning object. Techniques such as text clustering (Niemann et al., 2012), neural network classifiers (Roy, Sarkar, & Ghose, 2008), and collaborative tagging (Bateman, Brooks, McCalla, & Brusilovsky, 2007) have been used successfully to group, organize, and annotate different learning objects. More recently, with increased use of multimedia in education, different content analytics techniques have been used to automatically find important moments in lecture recordings to enhance navigation and use of video resources (Brooks, Amundson, & Greer, 2009; Brooks, Johnston, Thompson, & Greer, 2013).

In addition to organization and recommendation of learning resources, content analytics has been used to assess the quality of available instructional materials and how they impact learning outcomes. Dufty, Graesser, Louwerse, and McNamara (2006) showed that cohesiveness of the written text, as calculated by the Coh-metrix tool (Graesser, McNamara, & Kulikowich, 2011; McNamara, Graesser, McCarthy, & Cai, 2014), can successfully be used to evaluate the grade-level of textbooks, giving significantly better results than the simple text readability measures (e.g., Flesch Reading Ease, Flesch–Kincaid Grade Level, Degrees of Reading Power). Research has also revealed the direct link between the coherence of the provided learning materials and student comprehension of the subject domain (McNamara, Kintsch, Songer, & Kintsch, 1996; Varner, Jackson, Snow, & McNamara, 2013). The relationship between coherence and comprehension is also moderated by the students' level of background knowledge (Wolfe et al., 1998), which should be taken into account for recommending learning materials.

Content Analytics of Students' Products of Learning

One of the core goals of learning analytics is to enable provision of timely and relevant feedback to learners while studying (Siemens et al., 2011). One of the earliest domains where content analytics has been applied is the analysis of student essays, also known as automated essay scoring (AES). The most widely applied technique for automated essay scoring is Latent Semantic Analysis (LSA) (Landauer, Foltz, & Laham, 1998), used to measure the semantic similarity between two bodies of text through the analysis of their word co-occurrences. In the case of AES, LSA similarity is used to calculate the resemblance of an essay to a predefined set of essays and, based on those similarities, calculate a single, numeric measure of essay quality. In addition to LSA-based measures of essay quality, more recent systems such as WriteToLearn (Foltz & Rosenstein, 2015) include an extensive set of visualizations to provide students with feedback designed to help them acquire essay writing skills. While AES systems have been primarily used for the provision of real-time feedback (Crossley, Allen, Snow, & McNamara, 2015; Foltz et al., 1999; Foltz & Rosenstein, 2015), they could also be used for the (partial) automation of essay grading (Foltz et al., 1999), as they have shown to be as reliable and consistent as human graders.

Besides calculating the similarity of a text to a predefined collection of documents, LSA can also be used for calculating *internal* document similarity, often referred to as document coherence (the more coherent the document, the more semantically similar are its sentences). LSA is the underlying principle behind the Coh-metrix tool (Graesser et al., 2011; McNamara et al., 2014), often used to measure the quality of document writing. Coh-metrix has been extensively utilized for the analysis of different forms of written materials, including essays, discussion messages, and textbooks (McNamara et al., 2014). For example, it was adopted in Writing-Pal (McNamara et al., 2012), which is an intelligent tutoring system that provides students with feedback during essay writing exercises, looking at the essay's cohesiveness (calculated by Coh-metrix) as the indicator of its quality.

Another commonly adopted technique for the assessment of student essays are graph-based visualization methods, also based on a text's word co-occurrences. In addition to assessing the quality of writing, these tools are also used for summarizing essay content. For example, the OpenEssayist system (Whitelock, Field, Pulman, Richardson, & Van Labeke, 2014; Whitelock, Twiner, Richardson, Field, & Pulman, 2015) provides a graph-based overview of a student's essay in order to help the student visualize the relationship between different parts of the essay with the goal of teaching students how to write high-quality essays with a solid structure and a coherent narrative. Graph-based methods are also adopted for automated extraction of concept maps from students' collaborative writings. Such concept maps are then used to provide visual feedback to learners (Hecking & Hoppe, 2015) as a means of helping them review and revise their essays.

Besides approaches based on word co-occurrences, natural language processing techniques have also been used, in particular for the linguistic and rhetorical analysis of student essays. For instance, XIP Dashboard (Simsek, Buckingham Shum, De Liddo, Ferguson, & Sándor, 2014; Simsek, Buckingham Shum, Sandor, De Liddo, & Ferguson, 2013) visualizes meta-discourse of essays and highlights rhetorical moves and functions that help assess the quality of an argument in the essay (Simsek et al., 2014). These approaches to content analytics are also very similar to discourse-centric learning analytics (Buckingham Shum et al., 2013; Knight & Littleton, 2015) given that they use the same set of techniques for understanding the linguistic functions of the different parts of written text.

In addition to analyzing student essays, similar content analytics methods have been used for other types of student writing, most notably short answers (Burrows, Gurevych, & Stein, 2014). In the context of teaching physics, Dzikovska, Steinhauser, Farrow, Moore, and Campbell (2014) built a novel adaptive feedback system that takes into account the content of students' short answers, thus providing contextually relevant feedback. Likewise, the WriteEval system (Leeman-Munk, Wiebe, & Lester, 2014) evaluates students' short answers and provides feedback with follow-up instructions and tasks. As with essay grading, a set of reference answers facilitates the work of this group of systems. Similar approaches are also used for teaching troubleshooting skills (Di Eugenio, Fossati, Haller, Yu, & Glass, 2008), logic (Stamper, Barnes, & Croy, 2010), and PHP programming (Weragama & Reye, 2014). There have also been studies (Ramachandran, Cheng, & Foltz, 2015; Ramachandran & Foltz, 2015) showing the potential of using graph-based techniques for automated discovery of reference answers.

We should also note that many of the content analytics feedback systems have specifically been designed to provide instructors with feedback on student learning activities. For example, Lárusson and White (2012) used visualizations of student essays to inform instructors about the originality in student writings and particular points in time when students start to develop critical thinking. Besides providing feedback to students, automatic extraction of concept maps from student essays was also used to provide instructors with a broad overview of student learning activities (Pérez-Marín & Pascual-Nieto, 2010). Extraction of concept maps was also used for analysis of student chat logs (Rosen, Miagkikh, & Suthers, 2011), which are then used to provide instructors with an overview of social interactions and knowledge building among groups of students. Similarly, types of feedback and their effects on student engagement have also been explored. For instance, Crossley, Varner, Roscoe, and

McNamara (2013) investigated which types of feedback result in the biggest improvement in quality of student writing (based on the Coh-metrix analysis of student essays) while Calvo, Aditomo, Southavilay, and Yacef (2012) investigated how different types of feedback (i.e., directive, reflective) affect student essay editing behaviour. The ways in which students view and annotate video recordings has also been investigated (Gašević, Mirriahi, & Dawson, 2014; Mirriahi & Dawson, 2013) showing the potential for combining the analysis of different types of learning content.

A large body of work has also examined the association between different qualities of student essays and performance. The primary goal of these studies is to understand what encompasses successful writing (Allen, Snow, & McNamara, 2014; Crossley, Roscoe, & McNamara, 2014; McNamara, Crossley, & McCarthy, 2009; Snow, Allen, Jacovina, Perret, & McNamara, 2015), and how it relates to course performance (Robinson, Navea, & Ickes, 2013; Simsek et al., 2015). Current research has also revealed direct links between the coherence of the provided learning materials and the quality of students' reading summaries (Allen, Snow, & McNamara, 2015). Studies have also shown that insights into student comprehension of reading materials can be obtained through the analysis of their reading summaries using Coh-metrix cohesiveness measures and Information Content – a measure of text informativeness (Mintz, Stefanescu, Feng, D'Mello, & Graesser, 2014). Content analytics has also been used for understanding collaborative writing processes by using techniques such as Hidden Markov Models (Southavilay, Yacef, & Calvo, 2009, 2010) and probabilistic topic modelling (e.g., LDA; Southavilay, Yacef, Reimann, & Calvo, 2013). The same techniques are applied to understand how students learn to program (Blikstein, 2011), and even to analyze transcripts of student interviews to assess their expertise (Worsley & Blikstein, 2011) and knowledge of a given domain (Sherin, 2012).

Content Analytics of Students' Social Interactions

In online and distance education, asynchronous online discussions represent one of the primary means of interaction among students, and between students and instructors (Anderson & Dron, 2012). As such, insights into the overall discussion activity and contributions of different students are two areas where content analytics have been successfully applied, often using methods similar to those used for analyzing learning materials (e.g., LSA, Coh-metrix). Using LSA and Social Network Analysis (SNA), Teplovs, Fujita, and Vatrapu (2011) developed a visual analytics system that provides students with an overview of student contributions to online discourse. In addition to SNA, Hever et al. (2007) have also used process mining in combination with content analytics to raise awareness and enable better moderation of online discussions. Through the classification of student discussion messages based on their contribution type, textual content, and relationships (i.e., links) Hever et al. (2007) developed a message classification system that can be used to label discussion messages based on predefined theoretical or pedagogical categories. In addition to online discussions, raising instructor awareness of student activities in social media is explored by the LARAe system (Charleer, Santos, Klerkx, & Duval, 2014) showing the huge potential of social media for understanding student activities and learning progress. LARAe can automatically gather student social media postings (using RSS and Twitter API technologies) and then automatically assign one of 51 different badges to students, based on the observed social media activity. Instructors are then shown the collected information in the form of a dashboard for an easy overview of student activity and its change over time.

Online discussions have also been the focus of education researchers, who typically have used manual content analysis methods for parsing student discussion messages. Over the years, several content analytics systems have been developed to automate this process, in particular, analysis using the popular Community of Inquiry (CoI) framework (Garrison, Anderson, & Archer, 2001). For example, McKlin, Harmon, Evans, and Jones (2002) and McKlin (2004) developed a neural network classification system to automate coding of discussion messages for level of *cognitive presence*, the central construct of the CoI framework, focused on the development of students' critical and deep thinking skills. Building on results by McKlin (2004), a Bayesian network classification is used by the Automated Content Analysis Tool (Corich, Hunt, & Hunt, 2012) to provide a more generalizable model of classification that can be adopted for a wider range of coding schemes besides cognitive presence. More recently, several studies (Kovanović et al., 2014, 2016; Waters, 2015) examined the use of different text-mining techniques for coding messages for level of cognitive presence. Kovanović et al. (2014) developed a support vector machine classifier using different surface-level classification features (i.e., n-grams, part-of-speech n-grams, linguistic dependency triplets, the number of mentioned concepts, and discussion position metrics), which achieved higher classification accuracy than previous reports (McKlin, 2004; McKlin et al., 2002). The study by Waters (2015) also showed the benefits of using the structure of online discussions for text classification using conditional random fields, a structured classification technique that takes into the account relationships (i.e., reply-to structure) among individual classification instances (i.e., discussion messages).

Finally, a study by Kovanović et al. (2016) showed that metrics provided by the Coh-metrix (Graesser et al., 2011) and Linguistic Inquiry and Word Count (LIWC) tools (Tausczik & Pennebaker, 2010) – in combination with some of the NLP and discussion-position features – can be successfully used to develop a classification system almost as accurate as human coders. While further improvements are needed before this system can be widely adopted by educational researchers, the progress is promising and has the potential to advance research practices in content analysis.

With the social-constructivist view of learning and knowledge creation, a large body of work has utilized content analytics for understanding the role of social interactions on knowledge construction. For example, there has been significant research on linguistic differences – as captured by LIWC metrics – in discussion contributions (Joksimović, Gašević, Kovanović, Adesope, & Hatala, 2014; Xu, Murray, Park Woolf, & Smith, 2013) and how those differences relate to student grades (Yoo & Kim, 2012). Similarly, Chiu and Fujita (2014a, 2014b), investigated interdependencies between different types of discussion contributions with statistical discourse analysis (SDA), a group of statistical methods used to provide realistic modelling of student discourse interactions, while Yang, Wen, and Rosé (2014) used LDA and mixed membership stochastic blockmodels (MMSB) to examine what types of student discussion contributions are likely to receive response. Finally, using simple word frequency analysis, Cui and Wise (2015) examined what kinds of contributions are most likely to be acknowledged and answered by instructors. These and similar studies have the goal of understanding how interactions in online discourse eventually shape the learning outcomes and knowledge building. Similarly, different content analytics methods (text classification, topic modelling, mixed membership stochastic blockmodels) and tools (Coh-metrix, LIWC) have been applied to the products of student social interactions to gain a better understanding of students' (co-)construction of knowledge. These include research on the formation of student sub-communities (Yang, Wen, Kumar, Xing, & Rosé, 2014), development of self-regulation skills (Petrushyna, Kravcik, & Klamma, 2011), small-group communication (Yoo & Kim, 2013), and collaboration on computer programming projects (Velasquez et al., 2014). Further studies also investigated the link between accumulation of students' social capital in MOOCs (Dowell et al., 2015; Joksimović, Dowell et al., 2015; Joksimović, Kovanović et al., 2015), showing that position within the social network, extracted from learner interaction within various learning platforms, is associated with higher levels of cohesiveness of social media postings.

Content analytics has also been used extensively to assess the level of student engagement and instructional approaches that can contribute to its development. With this in mind, the analysis of student discussion messages – using a variety of content analytics methods – has commonly been used to assess the level of course engagement (Ramesh, Goldwasser, Huang, Daumé, & Getoor, 2013; Vega, Feng, Lehman, Graesser, & D'Mello, 2013; Wen, Yang, & Rosé, 2014b). Using probabilistic soft logic on both discussion content data and trace log data, Ramesh et al. (2013) examined student engagement in the MOOC context, focusing on the types of learners based on their level of discussion activity and course performance. Similarly, Wen, Yang, and Rosé (2014a) conducted a student sentiment analysis of MOOC online discussions, which revealed a strong association between expressed negative sentiment and the likelihood of dropping out of the course. Similar results are presented by Wen et al. (2014b) who also showed that LIWC word categories (most directly, cognitive words, first person pronouns, and positive words) could be used to measure the level of student motivation and cognitive engagement. Finally, by looking at the discrepancy between student reading time and text complexity, Vega et al. (2013) developed a content analytics system that can detect disengaged students. The general idea of using text complexity to measure engagement is that the easier the text, the shorter the reading time, unless the student is disengaged. This and similar types of analysis that combine trace data (e.g., text reading time) with the analysis of learning materials (e.g., analysis of text resource reading complexity) can be successfully used to monitor student motivation and engagement in real time, which is especially important for courses with large numbers of students, such as MOOCs.

Topic discovery in learning content

With huge amounts of web and other forms of learning data being available, one of the principal uses of content analysis is the organization and summarization of vast quantities of available information. In this regard, the most popular content analytics technique is probabilistic topic modelling, a group of methods used to identify key topics and themes in the collection of documents (e.g., discussion messages or social media posts). The most widely used topic modelling technique is latent Dirichlet allocation (LDA; Blei, 2012; Blei, Ng, & Jordan, 2003), which is often adopted in social sciences (Ramage, Rosen, Chuang, Manning, & McFarland, 2009) and digital humanities (Cohen et al., 2012). The general goal of LDA and other topic modelling techniques is to identify groups of words that are often used together, and which denote popular topics and themes in the document collection. Alongside LDA, techniques based on logic programming, text clustering, and LSA have also been used to extract main themes from student online discussions and social media postings.

Identification of main themes and topics has been extensively conducted in asynchronous online discussions. The primary goal is to raise instructors' awareness of the quality of student discourse by identifying the main themes and their magnitude in online discussions. For example, Antonelli and Sapino (2005) adopted a rule-based approach to modelling online discussions while the use of LDA has been explored by Chen (2014) and Hsiao and Awasthi (2015). In addition to topic modelling in online courses, given the large volume of discussions in massive open online courses (MOOCs), there has been particular interest in topic extraction from MOOC discussions using various approaches. Reich, Tingley, Leder-Luis, Roberts, and Stewart (2014) used structural topic models — an extension of the LDA technique that enables examining the differences in topics across different covariates — to investigate topics in MOOC online discussions and how different student (e.g., age, gender) and post characteristics (e.g., receiving an up-vote) relate to the identified topics. Likewise, Ezen-Can, Boyer, Kellogg, and Booth (2015) identified main themes in MOOC discussions through clustering "bag-of-words" representations of student online discussions.

While the discovery of topics in online discussions has been largely investigated, the analysis of main themes across different social media has received much less attention. One example is a study by Pham, Derntl, Cao, and Klamma (2012) who used SNA and word frequency analysis to investigate learning as it is occurring on popular blogging platforms and most important topics of discussion. In most of the studies, the focus of topic modelling analysis was primarily on traditional blogging platforms, while the analysis of micro-blogging platforms (e.g., Twitter) has received much less attention. In most cases, the reason for focusing on traditional blogging platforms is that most of the methods for topic modelling (e.g., LDA) are designed to work on longer text documents from which a correct topical distribution can be extracted (Zhao et al., 2011). Although several variations of LDA for short texts have been proposed (Hong & Davison, 2010; Mehrotra, Sanner, Buntine, & Xie, 2013; Ramage, Dumais, & Liebling, 2010; Yan, Guo, Lan, & Cheng, 2013), they are not currently widely used in the learning analytics field and their value is yet to be evaluated. One notable exception is the study by Chen, Chen, and Xing (2015) who — using ordinary LDA and SNA — analyzed tweets from the first four Learning Analytics and Knowledge conferences (LAK'11–LAK'14) and examined popular topics, as well as the structure and evolution of the learning analytics community over time. Similarly, a study by Joksimović, Kovanović et al. (2015) investigated the alignment between course materials and student postings in different social media (i.e., Facebook, Twitter, blogs). This study did not utilize traditional topic modelling techniques, but rather used a novel document clustering technique for topic discovery. Finally, topic modelling use has also been explored outside of social media. For example, a study by Reich et al. (2014) used LDA to examine major themes of student course evaluations, potentially providing an efficient, broad overview of course evaluation comments.

CONCLUSIONS AND FUTURE DIRECTIONS

In this chapter, we presented an overview of content analytics, a set of analytical methods and techniques for analyzing different forms of learning content in order to understand or improve learning activities. The wide range of research studies illustrates the great potential for applying content analytics techniques in addressing open problems in contemporary educational research and practice. In general, content analytics has been used for the analysis of 1) course resources, 2) student products of learning, and 3) student social interactions. Content analytics has been utilized to address a broad range of problems, such as recommendation and categorization of different learning materials (e.g., Drachsler et al., 2008), provision of feedback during student writing (e.g., Crossley et al., 2015), analysis of learning outcomes (e.g., Robinson et al., 2013), analysis of student engagement (e.g., Wen et al., 2014b), and topic discovery in online discussions (e.g., Reich et al., 2014). Given that learning analytics, as a research field, is still in its infancy, the list of problems being addressed by content analytics will likely expand in future. Likewise, as the field of content analytics matures, an important set of research practices and traditions will be established. Therefore, it is necessary to look toward future directions to provide the highest impact on educational research and practice. As such, we argue that current research in content analytics would be improved by 1) combining content analytics with other forms of analytics, and 2) developing content analytics systems based on existing educational theories. The early steps regarding the synergy between content analytics and other forms of analytics have already been observed. Several studies showed how content analytics could be successfully combined with

- **Discourse analytics** (Simsek et al., 2015, 2014, 2013),
- **Process mining** (Hever et al., 2007; Southavilay et al., 2009, 2010, 2013),
- **Social network analysis** (Drachsler et al., 2008; Joksimović, Kovanović et al., 2015; Joksimović et

al., 2014; Pham et al., 2012; Ramachandran & Foltz, 2015; Rosen et al., 2011; Teplovs et al., 2011),

- **Visual learning analytics** (Hecking & Hoppe, 2015; Lárusson & White, 2012; Pérez-Marín & Pascual-Nieto, 2010; Simsek et al., 2014; Whitelock et al., 2014, 2015), and
- **Multimodal learning analytics** (Blikstein, 2011; Worsley & Blikstein, 2011).

Likewise, it is important that additional forms of data — such as student demographics, prior knowledge, or standardized scores — are combined with content analytics, and in this regard, we also see some first steps (Crossley et al., 2015). Similar combined uses of traditional content analysis and other methods have been observed in mainstream online education research; more specifically, the use of social network analysis (De Laat, Lally, Lipponen, & Simons, 2007; Shea et al., 2010).

Finally, the development of content analytics should be based on well-established instructional theories. Many current approaches do not make use of the large body of educational research, which can limit the usefulness of the developed analytics systems and potentially even promote some detrimental learning practices (Gašević et al., 2015). Pedagogical considerations are particularly important for the provision of feedback, where the large body of previous research (Hattie & Timperley, 2007) demonstrates substantial differences in effectiveness between types of feedback provided. For example, the majority of feedback given by the current automated grading systems is summative in nature, although the most valuable feedback is on the process level, giving detailed instructions on identified weaknesses and suggestions for overcoming them. By building on existing educational knowledge, content analytics systems would not only increase in usefulness, but could also provide valuable opportunities for validation and refinement of the current understanding of learning processes.

REFERENCES

Allen, L. K., Snow, E. L., & McNamara, D. S. (2014). The long and winding road: Investigating the differential writing patterns of high and low skilled writers. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 304–307). International Educational Data Mining Society.

Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Are you reading my mind? Modeling students' reading comprehension skills with natural language processing techniques. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 246–254). New York: ACM. doi:10.1145/2723576.2723617

Anderson, T., & Dron, J. (2012). Learning technology through three generations of technology enhanced distance education pedagogy. *European Journal of Open, Distance and E-Learning, 2012*(II), 1–14.

Antonelli, F., & Sapino, M. L. (2005). A rule based approach to message board topics classification. In K. S. Candan & A. Celentano (Eds.), *Advances in Multimedia Information Systems* (pp. 33–48). Springer. http://link.springer.com/chapter/10.1007/11551898_6

Bateman, S., Brooks, C., McCalla, G., & Brusilovsky, P. (2007). Applying collaborative tagging to e-learning. Workshop held at the 16th International World Wide Web Conference (WWW2007), 8–12 May 2007, Banff, AB, Canada. http://www.www2007.org/workshops/paper_56.pdf

Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77–84. doi:10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Blikstein, P. (2011). Using learning analytics to assess students' behavior in open-ended programming tasks. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 110–116). New York: ACM. doi:10.1145/2090116.2090132

Bosnić, I., Verbert, K., & Duval, E. (2010). Automatic keywords extraction: A basis for content recommendation. *Proceedings of the 4th International Workshop on Search and Exchange of e-le@rning Materials* (SE@M'10), 27–28 September 2010, Barcelona, Spain (pp. 51–60). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.204.5641&rep=rep1&type=pdf#page=54

Bramucci, R., & Gaston, J. (2012). Sherpa: Increasing student success with a recommendation engine. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 82–83). New York: ACM. doi:10.1145/2330601.2330625

Brooks, C., Amundson, K., & Greer, J. (2009). Detecting significant events in lecture video using supervised machine learning. *Proceedings of the 2009 Conference on Artificial Intelligence in Education: Building Learning Systems That Care: From Knowledge Representation to Affective Modelling* (pp. 483–490). Amsterdam, The Netherlands: IOS Press. http://dl.acm.org/citation.cfm?id=1659450.1659523

Brooks, C., Johnston, G. S., Thompson, C., & Greer, J. (2013). Detecting and categorizing indices in lecture video using supervised machine learning. In O. R. Zaïane & S. Zilles (Eds.), *Advances in Artificial Intelligence* (pp. 241–247). Springer. http://link.springer.com/chapter/10.1007/978-3-642-38457-8_22

Buckingham Shum, S., De Laat, M. F., De Liddo, A., Ferguson, R., Kirschner, P., Ravenscroft, A., … Whitelock, D. (2013). DCLA13: 1st International Workshop on Discourse-Centric Learning Analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 282–282). New York: ACM. doi:10.1145/2460296.2460357

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Journal of Educational Technology & Society*, *15*(3), 3–26.

Buckingham Shum, S., Knight, S., McNamara, D., Allen, L., Bektik, D., & Crossley, S. (2016). Critical perspectives on writing analytics. *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 481–483). New York ACM. doi:10.1145/2883851.2883854

Burrows, S., Gurevych, I., & Stein, B. (2014). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, *25*(1), 60–117. doi:10.1007/s40593-014-0026-8

Calvo, R., Aditomo, A., Southavilay, V., & Yacef, K. (2012). The use of text and process mining techniques to study the impact of feedback on students' writing processes. *Proceedings of the 10th International Conference of the Learning Sciences* (ICLS '12), 2–6 July 2012, Sydney, Australia (pp. 416–423).

Cardinaels, K., Meire, M., & Duval, E. (2005). Automating metadata generation: The simple indexing interface. *Proceedings of the 14th International Conference on World Wide Web* (WWW '05), 10–14 May 2005, Chiba, Japan (pp. 548–556). ACM. http://dl.acm.org/citation.cfm?id=1060825

Charleer, S., Santos, J. L., Klerkx, J., & Duval, E. (2014). Improving teacher awareness through activity, badge and content visualizations. In Y. Cao, T. Väljataga, J. K..T. Tang, H. Leung, M. Laanpere (Eds.), *New Horizons in Web Based Learning* (pp. 143–152). Springer. http://link.springer.com/chapter/10.1007/978-3-319-13296-9_16

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, 4(5/6), 318–331. doi:10.1504/IJTEL.2012.051815

Chen, B. (2014). Visualizing semantic space of online discourse: The Knowledge Forum case. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 271–272). New York: ACM. doi:10.1145/2567574.2567595

Chen, B., Chen, X., & Xing, W. (2015). "Twitter archeology" of Learning Analytics and Knowledge conferences. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 340–349). New York: ACM. doi:10.1145/2723576.2723584

Chiu, M. M., & Fujita, N. (2014a). Statistical discourse analysis: A method for modeling online discussion processes. *Journal of Learning Analytics*, *1*(3), 61–83.

Chiu, M. M., & Fujita, N. (2014b). Statistical discourse analysis of online discussions: Informal cognition, social metacognition and knowledge creation. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 217–225). New York: ACM. doi:10.1145/2567574.2567580

Cohen, D. J., Troyano, J. F., Hoffman, S., Wieringa, J., Meeks, E., & Weingart, S. (Eds.). (2012). Special Issue on Topic Modeling in Digital Humanities. *Journal of Digital Humanities*, 2(1).

Cook, D. A., Garside, S., Levinson, A. J., Dupras, D. M., & Montori, V. M. (2010). What do we mean by web-based learning? A systematic review of the variability of interventions. *Medical Education*, 44(8), 765–774. doi:10.1111/j.1365-2923.2010.03723.x

Corich, S., Hunt, K., & Hunt, L. (2012). Computerised content analysis for measuring critical thinking within discussion forums. *Journal of E-Learning and Knowledge Society*, 2(1), 47–60.

Crossley, S., Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Pssst... Textual features... There is more to automatic essay scoring than just you! *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 203–207). New York: ACM. doi:10.1145/2723576.2723595

Crossley, S., Roscoe, R., & McNamara, D. S. (2014). What is successful writing? An investigation into the multiple ways writers can write successful essays. *Written Communication*, 31(2), 184–214. doi:10.1177/0741088314526354

Crossley, S., Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 269–278). Springer. http://link.springer.com/chapter/10.1007/978-3-642-39112-5_28

Cui, Y., & Wise, A. F. (2015). Identifying content-related threads in MOOC discussion forums. *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 299–303). New York: ACM. doi:10.1145/2724660.2728679

De Freitas, S. (2007). Post-16 e-learning content production: A synthesis of the literature. *British Journal of Educational Technology*, 38(2), 349–364. doi:10.1111/j.1467-8535.2006.00632.x

De Laat, M. F., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for Social Network Analysis. *International Journal of Computer-Supported Collaborative Learning*, 2(1), 87–103. doi:10.1007/s11412-007-9006-4

De Wever, B., Schellens, T., Valcke, M., & Van Keer, H. (2006). Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review. *Computers & Education*, 46(1), 6–28.

Di Eugenio, B., Fossati, D., Haller, S., Yu, D., & Glass, M. (2008). Be brief, and they shall learn: Generating concise language feedback for a computer tutor. *International Journal of Artificial Intelligence in Education*, 18(4), 317–345.

Donnelly, R., & Gardner, J. (2011). Content analysis of computer conferencing transcripts. *Interactive Learning Environments*, 19(4), 303–315.

Dowell, N., Skrypnyk, O., Joksimović, S., Graesser, A. C., Dawson, S., Gašević, D., ... Kovanović, V. (2015). Modeling learners' social centrality and performance through language and discourse. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Education Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 250–258). International Educational Data Mining Society. http://www.educationaldatamining.org/EDM2015/uploads/papers/paper_211.pdf

Drachsler, H., Hummel, H. G. K., & Koper, R. (2008). Personal recommender systems for learners in lifelong learning networks: The requirements, techniques and model. *International Journal of Learning Technology*, 3(4), 404–423. doi:10.1504/IJLT.2008.019376

Dufty, D. F., Graesser, A. C., Louwerse, M., & McNamara, D. S. (2006). Assigning grade levels to textbooks: Is it just readability? In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (CogSci 2006), 26–29 July 2006, Vancouver, British Columbia, Canada (pp. 1251–1256). Austin, TX: Cognitive Science Society.

Dzikovska, M., Steinhauser, N., Farrow, E., Moore, J., & Campbell, G. (2014). BEETLE II: Deep natural language understanding and automatic feedback generation for intelligent tutoring in basic electricity and electronics. *International Journal of Artificial Intelligence in Education*, 24(3), 284–332. doi:10.1007/s40593-014-0017-9

Ellis, C. (2013). Broadening the scope and increasing the usefulness of learning analytics: The case for assessment analytics. *British Journal of Educational Technology*, 44(4), 662–664. doi:10.1111/bjet.12028

Ezen-Can, A., Boyer, K. E., Kellogg, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 146–150). New York: ACM. doi:10.1145/2723576.2723589

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6), 304. doi:10.1504/IJTEL.2012.051816

Ferguson, R., & Buckingham Shum, S. (2012). Social learning analytics: Five approaches. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 23–33). New York: ACM. doi:10.1145/2330601.2330616

Foltz, P. W., Laham, D., Landauer, T. K., Foltz, P. W., Laham, D., & Landauer, T. K. (1999). Automated essay scoring: Applications to educational technology. In B. Collis & R. Oliver (Eds.), *Proceedings of EdMedia: World Conference on Educational Media and Technology* 1999, 19–24 June 1999, Seattle, WA, USA (pp. 939–944). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/p/6607

Foltz, P. W., & Rosenstein, M. (2015). Analysis of a large-scale formative writing assessment system with automated feedback. *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 339–342). New York: ACM. doi:10.1145/2724660.2728688

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence, and computer conferencing in distance education. *American Journal of Distance Education*, 15(1), 7–23.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. doi:10.1007/s11528-014-0822-x

Gašević, D., Mirriahi, N., & Dawson, S. (2014). Analytics of the effects of video use and instruction to support reflective learning. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 123–132). New York: ACM. doi:10.1145/2567574.2567590

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234. doi:10.3102/0013189X11413260

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487

Hecking, T., & Hoppe, H. U. (2015). A network based approach for the visualization and analysis of collaboratively edited texts. *Proceedings of the Workshop on Visual Aspects of Learning Analytics* (VISLA '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. XXX–XXX). http://ceur-ws.org/Vol-1518/paper4.pdf

Hever, R., De Groot, R., De Laat, M., Harrer, A., Hoppe, U., McLaren, B. M., & Scheuer, O. (2007). Combining structural, process-oriented and textual elements to generate awareness indicators for graphical e-discussions. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of the 7th International Conference on Computer-Supported Collaborative Learning* (CSCL 2007), 16–21 July 2007, New Brunswick, NJ, USA (pp. 289–291). International Society of the Learning Sciences. http://dl.acm.org/citation.cfm?id=1599600.1599654

Hong, L., & Davison, B. D. (2010). Empirical study of topic modeling in Twitter. *Proceedings of the 1st Workshop on Social Media Analytics* (SOMA '10), 25–28 July 2010, Washington, DC, USA (pp. 80–88). New York: ACM. doi:10.1145/1964858.1964870

Hosseini, R., & Brusilovsky, P. (2014). Example-based problem solving support using concept analysis of programming content. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent Tutoring Systems* (pp. 683–685). Springer. http://link.springer.com/chapter/10.1007/978-3-319-07221-0_106

Hsiao, I.-H., & Awasthi, P. (2015). Topic facet modeling: Semantic visual analytics for online discussion forums. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 231–235). New York: ACM. doi:10.1145/2723576.2723613

Joksimović, S., Dowell, N., Skrypnyk, O., Kovanović, V., Gašević, D., Dawson, S., & Graesser, A. C. (2015). Exploring the accumulation of social capital in cMOOC through language and discourse. *Under Review*.

Joksimović, S., Gašević, D., Kovanović, V., Adesope, O., & Hatala, M. (2014). Psychological characteristics in cognitive presence of communities of inquiry: A linguistic analysis of online discussions. *The Internet and Higher Education*, 22, 1–10.

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015). What do cMOOC participants talk about in social media? A topic analysis of discourse in a cMOOC. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 156–165). New York: ACM.

Knight, S., & Littleton, K. (2015). Discourse centric learning analytics: Mapping the terrain. *Journal of Learning Analytics*, *2*(1), 185–209.

Kovanović, V., Joksimović, S., Gašević, D., & Hatala, M. (2014). Automated content analysis of online discussion transcripts. In K. Yacef & H. Drachsler (Eds.), *Proceedings of the Workshops at the LAK 2014 Conference* (LAK-WS 2014), 24–28 March 2014, IN, Indiana, USA. http://ceur-ws.org/Vol-1137/LA_machinelearning_submission_1.pdf

Kovanović, V., Joksimović, S., Waters, Z., Gašević, D., Kitto, K., Hatala, M., & Siemens, G. (2016). Towards automated content analysis of discussion transcripts: A cognitive presence case. *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 15–24). New York: ACM. doi:10.1145/2883851.2883950

Krippendorff, K. H. (2003). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage Publications.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes*, *25*(2–3), 259–284. doi:10.1080/01638539809545028

Lárusson, J. A., & White, B. (2012). Monitoring student progress through their written "point of originality". *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 212–221). New York: ACM. doi:10.1145/2330601.2330653

Leeman-Munk, S. P., Wiebe, E. N., & Lester, J. C. (2014). Assessing elementary students' science competency with text analytics. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 143–147). New York: ACM. doi:10.1145/2567574.2567620

Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2011). Recommender systems in technology enhanced learning. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender systems handbook* (pp. 387–415). Springer. http://link.springer.com/chapter/10.1007/978-0-387-85820-3_12

McKlin, T. (2004). *Analyzing cognitive presence in online courses using an artificial neural network*. Georgia State University, College of Education, Atlanta, GA, United States. https://pdfs.semanticscholar.org/d6af/c0073f2efc53bb2e46a0dd39a677027b1c3d.pdf

McKlin, T., Harmon, S., Evans, W., & Jones, M. (2002, March 21). Cognitive presence in web-based learning: A content analysis of students' online discussions. *IT Forum*, 60. https://pdfs.semanticscholar.org/037b/f466c1c2290924e0ba00eec14520c091b57e.pdf

McNamara, D. S., Crossley, S., & McCarthy, P. M. (2009). Linguistic features of writing quality. *Written Communication*, *27*(1), 57–86. doi:10.1177/0741088309351547

McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction*, *14*(1), 1–43. doi:10.1207/s1532690xci1401_1

McNamara, D. S., Raine, R., Roscoe, R., Crossley, S., Jackson, G. T., Dai, J., … others. (2012). The Writing-Pal: Natural language algorithms to support intelligent tutoring on writing strategies. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing: Identification, investigation and resolution* (pp. 298–311). Hershey, PA: IGI Global.

Mehrotra, R., Sanner, S., Buntine, W., & Xie, L. (2013). Improving LDA topic models for microblogs via tweet pooling and automatic labeling. *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '13), 28 July–1 August 2013, Dublin, Ireland (pp. 889–892). New York: ACM. doi:10.1145/2484028.2484166

Mintz, L., Stefanescu, D., Feng, S., D'Mello, S., & Graesser, A. (2014). Automatic assessment of student reading comprehension from short summaries. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK. International Educational Data Mining Society. http://www.educationaldatamining.org/conferences/index.php/EDM/2014/paper/view/1372/1338

Mirriahi, N., & Dawson, S. (2013). The pairing of lecture recording data with assessment scores: A method of discovering pedagogical impact. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 180–184). New York: ACM. doi:10.1145/2460296.2460331

Moore, M. G. (1989). Editorial: Three types of interaction. *American Journal of Distance Education*, 3(2), 1–7. doi:10.1080/08923648909526659

Muldner, K., & Conati, C. (2010). Scaffolding meta-cognitive skills for effective analogical problem solving via tailored example selection. *International Journal of Artificial Intelligence in Education*, 20(2), 99–136.

Niemann, K., Schmitz, H.-C., Kirschenmann, U., Wolpers, M., Schmidt, A., & Krones, T. (2012). Clustering by usage: Higher order co-occurrences of learning objects. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 238–247). New York: ACM. doi:10.1145/2330601.2330659

Pérez-Marín, D., & Pascual-Nieto, I. (2010). Showing automatically generated students' conceptual models to students and teachers. *International Journal of Artificial Intelligence in Education*, 20(1), 47–72.

Petrushyna, Z., Kravcik, M., & Klamma, R. (2011). Learning analytics for communities of lifelong learners: A forum case. *Proceedings of the 11th IEEE International Conference on Advanced Learning Technologies* (ICALT '11), 6–8 July 2011, Athens, GA, USA (pp. 609–610). IEEE. doi:10.1109/ICALT.2011.185

Pham, M. C., Derntl, M., Cao, Y., & Klamma, R. (2012). Learning analytics for learning blogospheres. In E. Popescu, Q. Li, R. Klamma, H. Leung, & M. Specht (Eds.), *Advances in Web-Based Learning: ICWL 2012* (pp. 258–267). Springer. http://link.springer.com/chapter/10.1007/978-3-642-33642-3_28

Ramachandran, L., Cheng, J., & Foltz, P. (2015). Identifying patterns for short answer scoring using graph-based lexico-semantic text matching. *Proceedings of the 10th Workshop on Innovative Use of NLP for Building Educational Applications* (NAACL-HLT 2015), 4 June 2015, Denver, CO, USA (pp. 97–106). http://www.aclweb.org/anthology/W15-0612

Ramachandran, L., & Foltz, P. (2015). Generating reference texts for short answer scoring using graph-based summarization. *Proceedings of the 10th Workshop on Innovative Use of NLP for Building Educational Applications* (NAACL-HLT 2015), 4 June 2015, Denver, CO, USA (pp. 207–212). http://www.aclweb.org/anthology/W15-0624

Ramage, D., Dumais, S. T., & Liebling, D. J. (2010). Characterizing microblogs with topic models. In W. W. Cohen & S. Gosling (Eds.), *Proceedings of the 4th International AAAI Conference on Weblogs and Social Media* (ICWSM '10) 23–26 May 2010, Washington, DC, USA (pp. XXX–XXX). Palo Alto, CA: AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM10/paper/view/1528/1846

Ramage, D., Rosen, E., Chuang, J., Manning, C. D., & McFarland, D. A. (2009). Topic modeling for the social sciences. *Workshop on Applications for Topic Models: Text and Beyond* (NIPS 2009), 11 December 2009, Whistler, BC, Canada. https://ed.stanford.edu/sites/default/files/mcfarland/tmt-nips09-20091122+21-29-34.pdf

Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H., & Getoor, L. (2013). Modeling learner engagement in MOOCs using probabilistic soft logic. *NIPS Workshop on Data Driven Education* (NIPS-DDE 2013), 9 December 2013, Lake Tahoe, NV, USA. https://www.umiacs.umd.edu/~hal/docs/daume13engagementmooc.pdf

Reich, J., Tingley, D., Leder-Luis, J., Roberts, M. E., & Stewart, B. (2014). Computer-assisted reading and discovery for student generated text in massive open online courses. *Journal of Learning Analytics*, 2(1), 156–184.

Robinson, R. L., Navea, R., & Ickes, W. (2013). Predicting final course performance from students' written self-introductions: A LIWC analysis. *Journal of Language and Social Psychology*, 32(4). doi:10.1177/0261927X13476869

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 40(6), 601–618. doi:10.1109/TSMCC.2010.2053532

Rosen, D., Miagkikh, V., & Suthers, D. (2011). Social and semantic network analysis of chat logs. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 134–139). New York: ACM. doi:10.1145/2090116.2090137

Roy, D., Sarkar, S., & Ghose, S. (2008). Automatic extraction of pedagogic metadata from learning content. *International Journal of Artificial Intelligence in Education*, 18(2), 97–118.

Shea, P., Hayes, S., Vickers, J., Gozza-Cohen, M., Uzuner, S., Mehta, R., … Rangan, P. (2010). A re-examination of the community of inquiry framework: Social network and content analysis. *The Internet and Higher Education*, 13(1–2), 10–21. doi:10.1016/j.iheduc.2009.11.002

Sherin, B. (2012). Using computational methods to discover student science conceptions in interview data. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 188–197). New York: ACM. doi:10.1145/2330601.2330649

Siemens, G., Gašević, D., Haythornthwaite, C., Dawson, S., Buckingham Shum, S., Ferguson, R., … Baker, R. S. J. d. (2011, July 28). *Open learning analytics: An integrated & modularized platform.* SoLAR Concept Paper. http://www.elearnspace.org/blog/wp-content/uploads/2016/02/ProposalLearningAnalyticsModel_SoLAR.pdf

Simsek, D., Buckingham Shum, S., De Liddo, A., Ferguson, R., & Sándor, Á. (2014). Visual analytics of academic writing. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 265–266). New York: ACM. http://dl.acm.org/citation.cfm?id=2567577

Simsek, D., Buckingham Shum, S., Sandor, A., De Liddo, A., & Ferguson, R. (2013). XIP Dashboard: Visual analytics from automated rhetorical parsing of scientific metadiscourse. Presented at the 1st International Workshop on Discourse-Centric Learning Analytics, 8 April 2013, Leuven, Belgium. http://oro.open.ac.uk/37391/1/LAK13-DCLA-Simsek.pdf

Simsek, D., Sandor, A., Buckingham Shum, S., Ferguson, R., De Liddo, A., & Whitelock, D. (2015). Correlations between automated rhetorical analysis and tutors' grades on student essays. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 355–359). New York: ACM. http://dl.acm.org/citation.cfm?id=2723603

Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). You've got style: Detecting writing flexibility across time. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 194–202). New York: ACM. doi:10.1145/2723576.2723592

Southavilay, V., Yacef, K., & Calvo, R. A. (2009). WriteProc: A framework for exploring collaborative writing processes. *Proceedings of the 14th Australasian Document Computing Symposium* (ADCS 2009), 4 December 2009, Sydney, NSW, Australia (pp. 129–136). New York: ACM. http://es.csiro.au/adcs2009/proceedings/poster-presentation/09-southavilay.pdf

Southavilay, V., Yacef, K., & Calvo, R. A. (2010). Analysis of collaborative writing processes using hidden Markov models and semantic heuristics. In W. Fan, W. Hsu, G. I. Webb, B. Liu, C. Zhang, D. Gunopulos, & X. Wu (Eds.), *Proceedings of the 2010 IEEE International Conference on Data Mining Workshops* (ICDMW 2010), 14 December 2010, Sydney, Australia (pp. 543–548). doi:10.1109/ICDMW.2010.118

Southavilay, V., Yacef, K., Reimann, P., & Calvo, R. A. (2013). Analysis of collaborative writing processes using revision maps and probabilistic topic models. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 38–47). New York: ACM. doi:10.1145/2460296.2460307

Stamper, J., Barnes, T., & Croy, M. (2010). Enhancing the automatic generation of hints with expert seeding. In V. Aleven, J. Kay, & J. Mostow (Eds.), *Intelligent tutoring systems* (pp. 31–40). Springer. http://link.springer.com/chapter/10.1007/978-3-642-13437-1_4

Strijbos, J.-W., Martens, R. L., Prins, F. J., & Jochems, W. M. G. (2006). Content analysis: What are they talking about? *Computers & Education*, *46*(1), 29–48.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29. doi:10.1177/0261927X09351676

Teplovs, C., Fujita, N., & Vatrapu, R. (2011). Generating predictive models of learner community dynamics. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 147–152). New York: ACM. doi:10.1145/2090116.2090139

Varner, L. K., Jackson, G. T., Snow, E. L., & McNamara, D. S. (2013). Linguistic content analysis as a tool for improving adaptive instruction. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 692–695). Springer Berlin Heidelberg. doi:10.1007/978-3-642-39112-5_90

Vega, B., Feng, S., Lehman, B., Graesser, A., & D'Mello, S. (2013). Reading into the text: Investigating the influence of text complexity on cognitive engagement. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July, Memphis, TN, USA (pp. 296–299). International Educational Data Mining Society/Springer.

Velasquez, N. F., Fields, D. A., Olsen, D., Martin, T., Shepherd, M. C., Strommer, A., & Kafai, Y. B. (2014). Novice programmers talking about projects: What automated text analysis reveals about online scratch users' comments. *Proceedings of the 47th Hawaii International Conference on System Sciences* (HICSS-47), 6–9 January 2014, Waikoloa, HI, USA (pp. 1635–1644). IEEE Computer Society. doi:10.1109/HICSS.2014.209

Verbert, K., Manouselis, N., Ochoa, X., Wolpers, M., Drachsler, H., Bosnic, I., & Duval, E. (2012). Context-aware recommender systems for learning: A survey and future challenges. *IEEE Transactions on Learning Technologies*, 5(4), 318–335. doi:10.1109/TLT.2012.11

Walker, A., Recker, M. M., Lawless, K., & Wiley, D. (2004). Collaborative information filtering: A review and an educational application. *International Journal of Artificial Intelligence in Education*, 14(1), 3–28.

Waters, Z. (2015). *Using structural features to improve the automated detection of cognitive presence in online learning discussions* (B.Sc. Thesis). Queensland University of Technology.

Wen, M., Yang, D., & Rosé, C. (2014a). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK. International Educational Data Mining Society. http://www.cs.cmu.edu/~mwen/papers/edm2014-camera-ready.pdf

Wen, M., Yang, D., & Rosé, C. P. (2014b). Linguistic reflections of student engagement in massive open online courses. *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (ICWSM '14), 1–4 June 2014, Ann Arbor, Michigan, USA (pp. 525–534). Palo Alto, CA: AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8057

Weragama, D., & Reye, J. (2014). Analysing student programs in the PHP intelligent tutoring system. *International Journal of Artificial Intelligence in Education*, 24(2), 162–188. doi:10.1007/s40593-014-0014-z

Whitelock, D., Field, D., Pulman, S., Richardson, J. T. E., & Van Labeke, N. (2014). Designing and testing visual representations of draft essays for higher education students. 2nd International Workshop on Discourse-Centric Learning Analytics (DCLA14), 24 March 2014, Indianapolis, IN, USA. http://oro.open.ac.uk/41845/

Whitelock, D., Twiner, A., Richardson, J. T. E., Field, D., & Pulman, S. (2015). OpenEssayist: A Supply and demand learning analytics tool for drafting academic essays. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 208–212).

New York: ACM. doi:10.1145/2723576.2723599

Wolfe, M. B. W., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P. W., Kintsch, W., & Landauer, T. K. (1998). Learning from text: Matching readers and texts by latent semantic analysis. *Discourse Processes*, 25(2–3), 309–336. doi:10.1080/01638539809545030

Worsley, M., & Blikstein, P. (2011). What's an expert? Using learning analytics to identify emergent markers of expertise through automated speech, sentiment and sketch analysis. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero & J. Stamper (Eds.), *Proceedings of the 4th Annual Conference on Educational Data Mining* (EDM2011), 6–8 July 2011, Eindhoven, The Netherlands (pp. 235–240). International Educational Data Mining Society.

Xu, X., Murray, T., Park Woolf, B., & Smith, D. (2013). If you were me and I were you: Mining social deliberation in online communication. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July, Memphis, TN, USA (pp. 208–216). International Educational Data Mining Society/Springer.

Yan, X., Guo, J., Lan, Y., & Cheng, X. (2013). A biterm topic model for short texts. *Proceedings of the 22nd International Conference on World Wide Web* (WWW '13), 13–17 May 2013, Rio de Janeiro, Brazil (pp. 1445–1456). New York: ACM.

Yang, D., Wen, M., Kumar, A., Xing, E. P., & Rosé, C. P. (2014). Towards an integration of text and graph clustering methods as a lens for studying social interaction in MOOCs. *The International Review of Research in Open and Distributed Learning*, 15(5), 214–234.

Yang, D., Wen, M., & Rosé, C. (2014). Towards identifying the resolvability of threads in MOOCs. *Proceedings of the Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses at the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP 2014), 25 October 2014, Doha, Qatar (pp. 21–31). http://www.aclweb.org/anthology/W/W14/W14-41.pdf#page=28

Yoo, J., & Kim, J. (2012). Predicting learner's project performance with dialogue features in online Q&A discussions. In S. A. Cerri, W. J. Clancey, G. Papadourakis, & K. Panourgia (Eds.), *Intelligent tutoring systems* (pp. 570–575). Springer. http://link.springer.com/chapter/10.1007/978-3-642-30950-2_74

Yoo, J., & Kim, J. (2013). Can online discussion participation predict group project performance? Investigating the roles of linguistic features and participation patterns. *International Journal of Artificial Intelligence in Education*, 24(1), 8–32. doi:10.1007/s40593-013-0010-8

Zaldivar, V. A. R., García, R. M. C., Burgos, D., Kloos, C. D., & Pardo, A. (2011). Automatic discovery of complementary learning resources. In C. D. Kloos, D. Gillet, R. M. C. García, F. Wild, & M. Wolpers (Eds.), *Towards ubiquitous learning* (pp. 327–340). Springer. http://link.springer.com/chapter/10.1007/978-3-642-23985-4_26

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing Twitter and traditional media using topic models. In P. Clough, C. Foley, C. Gurrin, G. Jones, W. Kraaij, H. Lee, & V. Murdock (Eds.), *Proceedings of the 33rd European Conference on Advances in Information Retrieval* (ECIR 2011), 18–21 April 2011, Dublin, Ireland (pp. 338–349). Springer. http://dl.acm.org/citation.cfm?id=1996889.1996934

Chapter 8: Natural Language Processing and Learning Analytics

Danielle S. McNamara[1], Laura K. Allen[1], Scott A. Crossley[2], Mihai Dascalu[3], Cecile A. Perret[4]

[1]Psychology Department, Arizona State University, USA
[2]Applied Linguistics and ESL Department, Georgia State University, USA
[3]Computer Science Department, University Politehnica of Bucharest, Romania
[4]Institute for the Science of Teaching and Learning, Arizona State University, USA
DOI: 10.18608/hla17.008

ABSTRACT

Language is of central importance to the field of education because it is a conduit for communicating and understanding information. Therefore, researchers in the field of learning analytics can benefit from methods developed to analyze language both accurately and efficiently. Natural language processing (NLP) techniques can provide such an avenue. NLP techniques are used to provide computational analyses of different aspects of language as they relate to particular tasks. In this chapter, the authors discuss multiple, available NLP tools that can be harnessed to understand discourse, as well as some applications of these tools for education. A primary focus of these tools is the automated interpretation of human language input in order to drive interactions between humans and computers, or human–computer interaction. Thus, the tools measure a variety of linguistic features important for understanding text, including coherence, syntactic complexity, lexical diversity, and semantic similarity. The authors conclude the chapter with a discussion of computer-based learning environments that have employed NLP tools (i.e., ITS, MOOCs, and CSCL) and how such tools can be employed in future research.

Keywords: Natural language processing (NLP), language, computational linguistics, linguistic features, automated writing evaluation, intelligent tutoring systems, CSCL, MOOCs

Language is one means to externalize our thoughts. It allows us to express ourselves to others, to manipulate our world, and to label objects in the environment. Language allows us to internally construct and reconstruct our thoughts; it can represent our thoughts, and allow us to transform them. It allows us to construct and shape social experiences. Language provides a conduit to understanding and interacting with the world.

Language is omnipresent in our lives – in our thoughts, our communications, what we read and write, and our interactions with others. Language is equally central to education. Our goal as instructors is to communicate information to students so that they have the opportunity to learn new information, to absorb it, and to integrate it. Students are tasked with understanding language used to communicate information, and then to connect that information with what they already know – to construct their understanding as individuals, in groups, and in coordination with each other and instructors.

Language plays important roles in our lives, and in education, and thus, it is important to recognize and understand those roles and outcomes. Text and discourse analysis provides one means to understand complex processes associated with the use of language. Discourse analysts systematically examine structures and patterns within written text and spoken discourse and their relations to behaviours, psychological processes, cognition, and social interactions. Indeed, text and discourse analysis has provided a wealth of information about language.

Traditionally, however, discourse analysis is laborious. First, for example, the meaningful units of language are identified and segmented (e.g., clauses, utterances) and then experts code those units (i.e., with respect to the particular analysis). The potential relations between the nature of those language units and outcomes are then assessed. In a world of *big data*, where there are thousands of utterances and exchanges between individuals, hand-coding language is nearly impossible. Large corpora of data open the doors to understanding language on a wider and even more meaningful scale, but traditional approaches to discourse analysis are simply not feasible. One solution derives from *natural language processing* (NLP).

NLP is the analysis of human language using computers, providing the means to automate discourse analysis. The term NLP was coined because it is the analysis of natural human language, in contrast to the use and analysis of computer languages. A variety of automated tools can be used to process natural language. Indeed, the number and power of NLP tools have steadily increased since the mid-1990s (Jurafsky & Martin, 2000, 2008). As such, their impact and use within the realm of learning analytics and data mining is steadily, if not exponentially, increasing. This chapter describes several tools currently available to researchers and educators to analyze language computationally, focusing in particular on their uses in the realm of education.

NATURAL LANGUAGE PROCESSING

Computational linguistics is a discipline that focuses on the development of computational models of language. NLP tools and techniques are often guided by theories, models, and algorithms developed in the field of computational linguistics, but the primary purpose of NLP tools is the automated interpretation of human language input. Such an endeavor calls upon interdisciplinary perspectives integrating disciplines such as linguistics, computer science, psychology, and education. While NLP has a history dating back to Turing (1950), the majority of current NLP algorithms have been developed using a combination of NLP tools and data mining. A clear distinction must be made from the beginning between the NLP software often used by computer or data scientists and the tools presented in this chapter. A large majority of NLP research has focused on surface-level text processing (e.g., machine translation), and the available tools consequently emphasize the central role of accurate word- and sentence-level text processing. Our aim in this chapter is specifically to focus on NLP within the context of *learning analytics*. Thus, we focus on tools developed to calculate linguistic indices that move beyond these surface-level tasks and provide information that may be more important within educational contexts. Notably, we describe a subset of NLP techniques that provide information about multiple levels of text. These tools begin from the words in the discourse, extract specific word features, and then go beyond the lexicon by considering semantics, as well as discourse structure. Our goal is to provide examples of a few common techniques, rather than an overview of all available methods. We group these methods into those that focus on the words directly as the units of analysis, and those that focus on features of the words.

The Words

One approach to NLP is to analyze the words used in the language directly. For example, calculating the incidence of specific types of words within a text can reveal a good deal about the nature and purpose of the language used in various contexts. This is often referred to as a "bag-of-words" approach. One tool that employs this approach is the *Linguistic Inquiry Word Count* (LIWC) system developed by Pennebaker and colleagues (Pennebaker, Booth, & Francis, 2007; Pennebaker, Boyd, Jordan, & Blackburn, 2015; see http://liwc.wpengine.com/). The 2007 version of LIWC provides roughly 80 word categories, but also groups these word categories into broader dimensions. Examples of the broader dimensions are linguistic forms (e.g., pronouns, words in past tense, negations), social processes, affective processes, and cognitive processes. For example, cognitive processes include subcategories such as insight (e.g., think, know, consider), causation (e.g., because, effect, hence), and certainty (e.g., always, never). LIWC counts the number of words that belong to each word category and provides a proportion score that divides the number of words in the category by the total number of words in the text.

A similar approach is to identify n-grams, such as groups of characters or words, where *n* refers to the number of *grams* included in the group (e.g., bi-grams refer to groups of *two* words). N-gram analyses calculate probability distributions of word sequences in text and can provide information about the words common to a group of texts, or distinctive for a specific text or sets of texts (e.g., Jarvis et al., 2012). Several advantages of n-gram analyses include their simplicity and the potential for providing information about the specific content of a text, the linguistic and syntactic features of a text, and relationships between those features (Crossley & Louwerse, 2007).

The Features of the Words

Calculating the occurrence of words and groups of words considers the explicit content of the text. An alternative approach involves the calculation of the features of the words and sentences in a text. One such technique is to derive the latent meaning behind the words (McNamara, 2011). There are numerous algorithms for doing so, but the most well known and perhaps the first was Latent Semantic Analysis (LSA; Landauer & Dumais, 1997; Landauer, McNamara, Dennis, & Kintsch, 2007; see lsa.colorado.edu). LSA emerged in the mid-1990s, providing a means to extract semantic meaning from large bodies of text, and to compare large and small text samples for semantic similarities. Such an approach provided a unique potential to revolutionize NLP. LSA is a mathematical, statistical technique that uses *singular value decomposition* to compress (i.e., factorize) a matrix representing the occurrence of words across a large set of documents. A principal assumption driving LSA is that the meanings of words are captured by the company they keep. For example, the word "data" will be highly associated with words of the same functional context, such as "computations", "mining", "computer", and "mathematics". These words do not mean the same thing as *data*. Rather, these words are related to *data* because they typically occur in similar contexts. By affording the computation of the semantic similarities between words, sentences, and paragraphs, LSA opened the doors to the simulation of meaning in text (McNamara, 2011). LSA can be considered the first word-based approach to successfully address the question of relevance (i.e., the degree to which a text is relevant to another text or to a core concept), a problem for which simple measures of word overlap are not sufficient. While there are multiple approaches that have gone beyond LSA (see McNamara, 2011, for an overview), LSA remains a common approach used across multiple contexts to model word meaning and to provide insights in terms of semantics and text cohesion (e.g., Landauer et al., 2007; McNamara, Graesser, McCarthy, & Cai, 2014).

One obvious feature of language is the meaning, but many other features can be derived from linguistic analyses, such as the parts of speech (e.g., verb, noun), syntax, psychological aspects (e.g., concreteness, meaningfulness), and the relations between ideas in the text (e.g., cohesion). Coh-Metrix is an example of an automated language analysis tool, first launched in 2003, that uses multiple sources of information about language to extract linguistic, psychological, and semantic features of text (McNamara et al., 2014; cohmetrix.com). Coh-Metrix adapts and integrates information about the English language from a variety of sources including LSA, the *MRC Psycholinguistic Database*, *WordNet*, and word frequency indices such as CELEX, as well as syntactic parsers. For example, the MRC Psycholinguistic Database provides psycholinguistic information about words (Wilson, 1988) and WordNet provides linguistic and semantic features of words, as well as semantic relations between words (Miller, Beckwith, Fellbaum, Gross, & Miller, 1990). Coh-Metrix also calculates linguistic indices related to various aspects of language through simple features of text quality, such as word frequency and sentence length, as well as more complex features such as coherence and syntactic complexity, in order to produce a multi-dimensional analysis of written or spoken text (McNamara, Ozuru, Graesser, & Louwerse, 2006). Coh-Metrix can provide a relatively simple characterization of a text through descriptive indices (i.e., length of words, sentences, paragraphs). In addition, it offers various complex indices that describe a text's quality and readability. Among these indices are the five Coh-Metrix Text Easability Components, including narrativity, referential cohesion, syntactic simplicity, word concreteness, and deep cohesion (Graesser, McNamara, & Kulikowich, 2011; Jackson, Allen, & McNamara, 2016; see coh-metrix.commoncoretera.com).

Coh-Metrix has had a large impact on our understanding of language and discourse by making automated language analysis publicly available. While Coh-Metrix provides multiple measures of language, the primary, unique focus of Coh-Metrix has been on providing measures of cohesion in text. Cohesion is the overlap in features, words, and meaning between sentences (i.e., local cohesion) and larger sections of the text such as paragraphs (i.e., global cohesion) and the text overall (e.g., lexical diversity). While extremely useful, Coh-Metrix has had several shortcomings regarding facile and broad measurement of cohesion indices. First, it does not allow for the batch processing of text, and it is not housed on a user's hard drive (and thus it depends on an internet connection and an external server). Second, Coh-Metrix cohesion indices generally focus on local and overall text cohesion (i.e., average sentence overlap, lexical diversity), rather than global cohesion (e.g., semantic overlap between various sections of a text). Hence, the Tool for the Automatic Analysis of Text Cohesion (TAACO) and the Simple Natural Language Processing Tool (SiNLP) were developed to address these gaps (Crossley, Allen, Kyle, & McNamara, 2014; Crossley, Kyle, & McNamara, in press; http://www.kristopherkyle.com/taaco.html). TAACO is locally installed (as compared to an internet interface), allows for batch processing of text files, and includes over 150 indices related to local, global, and overall text cohesion. Similarly, SiNLP is locally installed and allows for batch text processing. However, SiNLP differs from TAACO in that it takes the "bag-of-words" approach to calculate information about multiple aspects of texts. Additionally, the tool is flexible and allows researchers to add their own categories of words to inform additional analyses.

Another example of a freely available NLP tool is the

Tool for the Automatic Analysis of LExical Sophistication (TAALES; Kyle & Crossley, 2015; http://www.kristopherkyle.com/taales.html). TAALES focuses on providing extensive information about the level of lexical sophistication present in a text. This type of analysis is important because it provides information on the lexical demands of a text, as well as potential information related to the lexical knowledge of the author of the text (Kyle & Crossley, 2015). TAALES calculates over 130 classic and newly developed lexical indices to assess the breadth and depth of lexical knowledge used in a text. This tool is fast, reliable, and freely available for download. The measures for TAALES include word frequency, word and word family range, n-grams, academic lists, and word information indices that consider psycholinguistic components (Kyle & Crossley, 2015). These indices collectively provide extensive information on the complexity of word choices in text.

Dascalu, McNamara, Crossley, and Trausan-Matu (2016) also introduced *Age of Exposure* (AoE), a computational model to estimate word complexity in which the learning rate of individual words is calculated as a function of a learner's experience with language. In contrast to Pearson's calculation of word maturity (Landauer, Kireyev, & Panaccione, 2011), AoE is a reproducible and scalable model that simulates word learning in terms of potential associations that can be created with it across time or, more specifically, across incremental latent Dirichlet allocation (Blei, Ng, & Jordan, 2003) topic models. AoE indices yield strong associations (exceeding the reported performance of word maturity) with estimates of word frequency and entropy, as well as human ratings of age of acquisition and lexical response latencies.

Natural Language Processing and Learning Algorithms

NLP can be used to describe multiple facets of language from simple descriptive statistics, such as the number of words, n-grams, and paragraphs, to the features of words, sentences, and text (Crossley, Allen, Kyle, & McNamara, 2014). As depicted in Figure 8.1, multiple characteristics of language can be gleaned from the words (including n-grams and bags of words) and captured using both techniques for analyzing observable features (e.g., word frequencies, word-document distributions) and latent meaning from the text (McNamara, 2011). Information is provided by the features of the words, the sentences, and the text as a whole. This information can be analyzed using machine learning techniques such as linear regression, discriminant function classifiers, Naïve-Bayes classifiers, support vector machines, logistic regression classifiers, and decision tree classifiers. When these techniques are used to predict learning outcomes, algorithms can be derived that can then be used within educational technologies or applications. We discuss a number of these applications in the following sections.

WRITING ASSESSMENT

The most common example of the use of NLP in the realm of education is for the development of automated essay scoring (AES) algorithms (Allen, Jacovina, & McNamara, 2016; Dikli, 2006; Weigle, 2013; Xi, 2010). AES

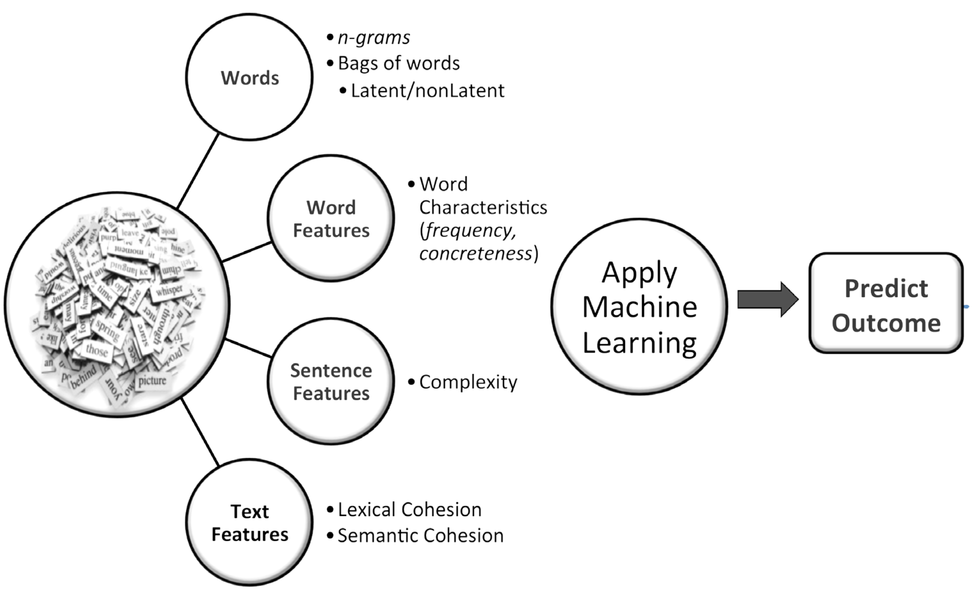


Figure 8.1. Developing algorithms using NLP requires machine-learning techniques applied to various sources of information on the text, including information from the words, sentences, and the entire text.

systems assess essays using a variety of approaches. For example, the Intelligent Essay Assessor (Landauer, Laham, & Foltz, 2003) primarily relies on LSA to assess the similarity of an essay to benchmark essays. By contrast, systems such as the e-rater developed at Educational Testing Service (Burstein, Chodorow, & Leacock, 2004), IntelliMetric Essay Scoring System developed by Vantage Learning (Rudner, Garcia, & Welch, 2006), and the Writing Pal (McNamara, Crossley, & Roscoe, 2013) rely on combinations of NLP techniques and artificial intelligence. AES systems process writing samples such as essays, and assess the degree to which the writer has met the demands of the task by assessing the quality of essays and their accuracy relative to the content. AES technologies are highly successful, reporting levels of accuracy generally as accurate as expert human raters (Attali & Burstein, 2006; Shermis, Burstein, Higgins, & Zechner, 2010; Valenti, Neri, & Cucchiarelli, 2003; Crossley, Kyle, & McNamara, 2015).

Tutoring Systems

Another use of NLP has been in the context of automated, intelligent tutoring technologies. NLP has been incorporated into a number of intelligent tutoring systems (ITSs), particularly those that interact with the student via dialogue (e.g., AutoTutor: Graesser et al., 2004) and those that prompt the student to generate verbal responses (e.g., iSTART: McNamara, Levinstein, & Boonthum, 2004; Writing Pal: McNamara et al., 2012; Roscoe & McNamara, 2013). When a student enters natural language into a system and expects useful feedback or a reasonable response, NLP can be used to interpret that input and provide appropriate feedback (McNamara et al., 2013). For tutoring systems that accept natural language as input (e.g., verbal explanations of text, problems, or scientific processes), student responses can be open-ended and potentially ambiguous. For example, the student might be asked which phase of cell mitosis involves the lengthening of the microtubules. This type of question (e.g., *what* or *when* questions) can be answered using short answers or multiple-choice responses, requiring little to no NLP. By contrast, a question to describe the process of Anaphase would elicit answers likely to differ widely between students. Thus, automatically detecting the accuracy and quality of the student's answer requires the use of NLP.

Why not just use multiple-choice? Many tutorial systems do just that. However, students are more likely to construct a deep understanding of a construct or phenomenon by answering *how* and *why* questions (e.g., Johnson-Glenberg, 2007; McKeown, Beck, & Blake, 2009; Wong, 1985). Moreover, students' answers to these types of questions are more likely to unveil the depth of their understanding (Graesser & Person, 1994; Graesser, McNamara, & VanLehn, 2005; McNamara & Kintsch, 1996). AutoTutor is an ITS that focuses on providing instruction on challenging topics (e.g., physics, biology, computer programing) by prompting students to answer deep level *how* and *why* questions. AutoTutor engages the student via an animated agent in a dialogue that moves the student toward constructing the correct answers. It does so by using a variety of dialogue moves, such as hints, prompts, assertions, corrections, and answers to student questions. These moves are driven by a combination of NLP techniques. For example, AutoTutor uses frozen expressions to detect phrases that students are likely to produce in certain situations (e.g., *I don't know*; *I don't understand*) as well as key parts of the correct answer. AutoTutor also uses LSA to detect the similarity between the answer provided by the student and the ideal answer. The combination of frozen expressions, regular expressions or patterns, inverse-frequency weighted word overlaps between student verbal responses and expectations, and LSA, allows AutoTutor to simulate the understanding of the student's answer, and in turn, this simulated understanding drives an appropriate response to the student (Graesser, in press).

iSTART (Interactive Strategy Training for Active Reading and Thinking) is another ITS that relies on a combination of NLP techniques to respond to open-ended responses. iSTART was among the first automated systems to address the paraphrase problem in student's self-explanations, a difficult challenge in the both NLP and computational linguistics literature. iSTART enhances students' comprehension of challenging science texts by providing instruction and practice to use self-explanation (i.e., the process of explaining text to oneself) in combination with comprehension strategies such as generating bridging and elaborative inferences. During the practice phase of iSTART instruction, students generate self-explanations for challenging texts. Students' self-explanations in iSTART are scored using an algorithm that combines information from the words in the self-explanation and the text, using a combination of observable and latent semantic information about the words (McNamara, Boonthum, Levinstein, & Millis, 2007). The algorithm automatically assigns a score between 0 and 3 to each self-explanation. Higher scores are assigned to self-explanations that include information related to the text content (both the target sentence and previously read sentences), whereas lower scores are assigned to unrelated or short responses. The scoring algorithm is designed to reflect the extent to which students construct connections between the target sentence, prior text content, and world knowledge. The system successfully matches human scores of the explanations across a wide variety of texts (Jackson, Guess, & McNamara, 2010; McNamara et al., 2007).

Computer Supported Collaborative Learning (CSCL)

NLP techniques have also been applied to discourse generated in collaborative learning environments, and in particular Computer Supported Collaborative Learning (CSCL) systems (Stahl, 2006). A subset of these systems model CSCL conversations based on *dialogism*, a concept introduced by Bakhtin (1981) that later on emerged as a paradigm for CSCL (Koschmann, 1999). The most representative approaches are Dong's (2005) design of team communication, Polyphony (Trausan-Matu, Rebedea, Dragan, & Alexandru, 2007), the Knowledge Space Visualizer (Teplovs, 2008), and ReaderBench (Dascalu, Stavarache et al., 2015; Dascalu, Trausan-Matu, McNamara, & Dessus, 2015). ReaderBench leverages the power of text mining techniques, advanced NLP, and social network analysis to achieve multiple objectives related to language comprehension as well as collaborative learning (Dascalu, 2014). ReaderBench models participation and collaboration from a Cohesion Network Analysis perspective in which the information communicated among participants is computed via semantic textual cohesion (Dascalu, Trausan-Matu, Dessus, & McNamara, 2015a). Moreover, ReaderBench has introduced an automated dialogic model for assessing collaboration based on the polyphonic model of discourse (Trausan-Matu, Stahl, & Sarmiento, 2007). Grounded in theories of dialogism (Bakhtin, 1981), the system automatically identifies *voices* or participant's points of view as semantic chains that include tightly cohesive or semantically related concepts spanning throughout the entire conversation (Dascalu, Trausan-Matu, Dessus, & McNamara, 2015b). Thus, collaboration emerges from the inter-animation of different participant voices, which is computationally captured in the co-occurrence patterns used to highlight the exchange of ideas between different participants.

Massive Open Online Courses (MOOCs)

Another use of NLP has been in the context of online courses, particularly massive open online courses (MOOCs). MOOCs use online platforms to make courses available to thousands of students without cost to the student. MOOCs are lauded for their potential to increase accessibility to distance and lifelong learners (Koller, Ng, Do, & Chen, 2013). These platforms can provide a tremendous amount of data via click-stream logs, assignments, course performance, as well as language generated by the students within discussion forums and emails. These data can be mined to examine student attitudes, completion, and learning (Seaton, Bergner, Chuang, Mitros, & Pritchard, 2014; Wen, Yang, & Rosé, 2014a, 2014b).

The most common NLP approach to analyzing student language in MOOCS has been through tools that analyze emotions. Sentiment analysis examines language for positive or negative emotion words or words related to motivation, agreement, cognitive mechanisms, or engagement (Chaturvedi, Goldwasser, & Daumé, 2014; Elouazizi, 2014; Moon, Potdar, & Martin, 2014; Wen et al., 2014a, 2014b). For example, Moon et al. (2014) used emotion terms and semantic similarity among participants to identify student leaders. Elouazizi (2014) showed that linguistic indices related to point of view (e.g., *think*, *believe*, *presumably*, *probably*) were correlated with low levels of engagement in the course. Wen and colleagues (2014a, 2014b) found that students' use of personal pronouns and words related to motivation within discussion forums was predictive of a lower risk of dropping out of the course.

Similarly, Crossley, McNamara et al. (2015) used multiple levels of linguistic features to examine students' language in a MOOC discussion forum within a course covering the topic of educational data mining (Baker et al., in press). Crossley, McNamara et al. (2015) successfully predicted the completion rates (with an accuracy of 70%) of 320 students who participated within the MOOC discussion forums (i.e., posted > 49 words). Students who were more likely to receive a certificate of completion in the course generally used more sophisticated language. For example, their posts were more concise and cohesive, used less frequent and specific words, and had greater overall writing quality. Interestingly, indices related to affect were not predictive of completion rates.

Collectively, this research provides promising evidence that NLP can be a powerful predictor of success in the context of MOOCs. Communication between the instructor and the students as well as between the students is crucial, particularly for distance courses. Further, this communication can then be used as forms of assessment of student performance. Therefore, it seems apparent that MOOCs should include discussion forums in order to better monitor student participation and potential success. The language that students use can also be utilized to identify students who are less likely to complete the course, and target those students for interventions such as sending emails, suggesting content, or recommending tutoring. Automating language understanding, and thereby providing information about the language and social interactions within these courses, will help to enhance both learning and engagement in MOOCs.

THE POWER OF NLP

NLP is extremely powerful, primarily because language is ubiquitous and also because tools to analyze language automatically provide indices related to virtually any aspect of language (Crossley, 2013). NLP can detect the specific words used, groups of words, and the strength of the relations between words and between larger bodies of text. It can also detect the features of the text, such as the frequency, concreteness, or meaningfulness of the words, the complexity of the sentences, and various aspects of the text such as cohesion and genre. The words and their features serve as proxies to various constructs. For example, the frequency of the words in a text serves as a proxy to estimate the knowledge that might be required to understand the text. The cohesion of a text affords an estimate of the knowledge necessary to fill in the gaps in a text.

NLP has been used to identify a wide variety of other constructs. For example, Crossley and McNamara (2012) demonstrated that the linguistic features of second language (L2) writers' essays could predict the native language of those writers. Varner, Roscoe, and McNamara (2013) used indices provided by both Coh-Metrix and LIWC to examine differences in students' and teachers' ratings of essay quality. Louwerse, McCarthy, McNamara, and Graesser (2004) used NLP techniques to identify differences between spoken and written samples of English. McCarthy, Briner, Rus, and McNamara (2007) showed that Coh-Metrix could differentiate sections in typical science texts, such as *introductions*, *methods*, *results*, and *discussions*. Additionally, Crossley, Louwerse, McCarthy, and McNamara's (2007) investigations of second language learner texts, revealed a wide variety of structural and lexical differences between texts that were *adopted* (or authentic) versus *adapted* (or simplified) for second language learning purposes. Finally, NLP has also been used to detect deception. Duran, Hall, McCarthy, and McNamara (2010) examined the extent to which features of language discriminated between conversational dialogues in which a person was being deceptive and those in which the person was being truthful.

It is important to note that there are potential drawbacks to using NLP. For example, certain NLP techniques rely on simplified representations of dialogue that use word counts or "bag-of-words" approaches. The most notable and widely used NLP word representations, including LSA vector-spaces, latent Dirichlet allocation topic distributions (LDA; Blei et al., 2003), and word2vec models based on neural networks (Mikolov, Chen, Corrado, & Dean, 2013), are all subject to the "bag-of-words" assumption in which word order is disregarded. In addition, many NLP analyses ignore context, such as the intentions or pragmatic aspects of the speaker. Similarly, NLP analyses are often limited to particular corpora and situations, and fail to generalize to other contexts. Even with these (and other) caveats, NLP is extremely powerful. Because of the vast sources of information now available from NLP tools, and because the language we use can be an extension or externalization that represents thoughts and intentions, NLP can provide information about the individuals, their abilities, their emotions, their intentions, and social interactions. In the context of learning analytics, it is a means toward the automated understanding of learning processes and the learner.

The Big Picture

NLP provides techniques that automate the analysis of language, which allows researchers to establish a better understanding of language and of the roles that language potentially plays in various aspects of our lives. NLP informs feedback systems within tu-

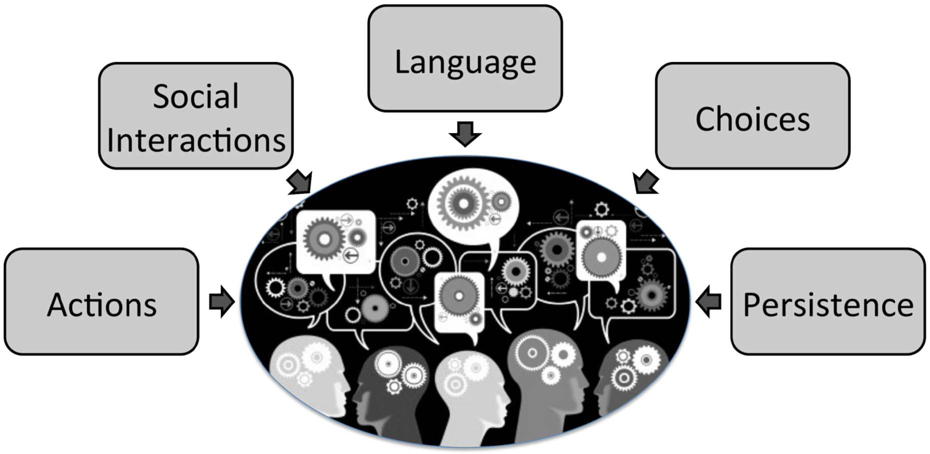


Figure 8.2. Predicting educational outcomes will require the integration of multiple sources of data.

toring systems that prompt the student to generate language within answers to questions, explanations, and essays. NLP provides a means of simulating intelligence within language-based tutoring systems. NLP is also informative in the context of online discussion forums. It provides information on student attitudes, motivation, and the quality of the language, which in turn is predictive of students' likelihood of performing well or completing the course.

One goal of learning analytics is to model the characteristics and skills of students in order to provide more effective instruction (Allen & McNamara, 2015). Specifically, we can use this data for various purposes: provide automated feedback on performance, intervene during learning, provide scaffolding or support, recommend tutoring, personalize learning, and so on, with the assumption that information gleaned from analytics will ultimately enhance learning. For this purpose, researchers are increasingly turning to large, complex data sources (i.e., big data) and using various combinations of data types and analytic techniques. NLP is crucial to this endeavour because the proposed techniques help to improve student learning through the prediction and assessment of comprehension across a variety of contexts. However, NLP is only one piece of the puzzle.

As depicted in Figure 8.2, developing a complete and highly predictive understanding of student outcomes requires multiple sources of information and a variety of approaches to data analysis. Learning is a complex process with multiple layers and multiple time scales. Relying on any single source or type of data to understand the learning process is myopic, particularly when so many automated sources of information are currently available. NLP is simply one source of data increasingly recognized as an integral piece of the big picture that ultimately we seek. Developing a complete understanding of learning will require an integration of multiple sources of data.

ACKNOWLEDGMENTS

The writing of this chapter was supported in part by the Institute for Educational Sciences (IES R305A120707, R305A130124), the National Science Foundation (NSF DRL-1319645, DRL-1418352, DRL-1418378, DRL-1417997), and the Office for Naval Research (ONR N000141410343). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES, NSF, or ONR. We are also grateful to the many students, postdoctoral researchers, and faculty who have contributed to our research on NLP over the years.

REFERENCES

Allen, L. K., Jacovina, M. E., & McNamara, D. S. (2016). Computer-based writing instruction. In C. A. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research*, 2nd ed. (pp. 316–329). New York: The Guilford Press.

Allen, L. K., & McNamara, D. S. (2015). You are your words: Modeling students' vocabulary knowledge with natural language processing. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM 2015) 26–29 June 2015, Madrid, Spain (pp. 258–265). International Educational Data Mining Society.

Attali, Y., & Burstein, J. (2006). Automated essay scoring with e-rater® V. 2. *The Journal of Technology, Learning and Assessment*, 4(2). doi:10.1002/j.2333-8504.2004.tb01972.x

Baker, R., Wang, E., Paquette, L., Aleven, V., Popescu, O., Sewall, J., Rose, C., Tomar, G., Ferschke, O., Hollands, F., Zhang, J., Cennamo, M., Ogden, S., Condit, T., Diaz, J., Crossley, S., McNamara, D., Comer, D., Lynch, C., Brown, R., Barnes, T., & Bergner, Y. (in press). A MOOC on educational data mining. In. S. ElAtia, O. Zaïane, & D. Ipperciel (Eds.). *Data Mining and Learning Analytics in Educational Research*. Wiley & Blackwell.

Bakhtin, M. M. (1981). *The dialogic imagination: Four essays* (C. Emerson & M. Holquist, Trans.). Austin, TX: University of Texas Press.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3(4–5), 993–1022.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion online writing service. *Ai Magazine*, 25(3), 27.

Chaturvedi, S., Goldwasser, D., & Daumé III, H. (2014). Predicting instructor's intervention in MOOC forums. In D. Marcu, K. Toutanova, & H. W. Baidu (Eds.), *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (pp. 1501–1511). Baltimore, MD.

Crossley, S. A. (2013). Advancing research in second language writing through computational tools and machine learning techniques: A research agenda. *Language Teaching*, 46(2), 256–271.

Crossley, S. A., Allen, L. K., Kyle, K., & McNamara, D. S. (2014). Analyzing discourse processing using a simple natural language processing tool (SiNLP). *Discourse Processes*, 51, 511–534.

Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). To aggregate or not? Linguistic features in automatic essay scoring and feedback systems. *Journal of Writing Assessment*, 8(1). http://www.journalofwritingassessment.org/article.php?article=80

Crossley, S. A. Kyle, K., & McNamara, D. S. (in press). Tool for the automatic analysis of text cohesion (TAACO): Automatic assessment of local, global, and text cohesion. *Behavior Research Methods*.

Crossley, S. A., & Louwerse, M. (2007). Multi-dimensional register classification using bigrams. *International Journal of Corpus Linguistics*, 12(4), 453–478.

Crossley, S. A., Louwerse, M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, 91, 15–30.

Crossley, S. A., & McNamara, D. S. (2012). Interlanguage talk: A computational analysis of non-native speakers' lexical production and exposure. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 425–437). Hershey, PA: IGI Global.

Crossley, S. A., McNamara, D. S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: Success in an educational data mining massive open online class. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM 2015), 26–29 June 2015, Madrid, Spain (pp. 388–391). International Educational Data Mining Society.

Dascalu, M. (2014). Analyzing discourse and text complexity for learning and collaborating. *Studies in Computational Intelligence* (Vol. 534). Switzerland: Springer.

Dascalu, M., Stavarache, L. L., Dessus, P., Trausan-Matu, S., McNamara, D. S., & Bianco, M. (2015). ReaderBench: The learning companion. In A. Mitrovic, F. Verdejo, C. Conati, & N. Heffernan (Eds.), *Proceedings of the 17th International Conference on Artificial Intelligence in Education* (AIED '15), 22–26 June 2015, Madrid, Spain (pp. 915–916). Springer.

Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara, D. S. (2015a). Discourse cohesion: A signature of collaboration. In P. Blikstein, A. Merceron, & G. Siemens (Eds.), *Proceedings of the 5th International Learning Analytics & Knowledge Conference* (LAK '15), 16–20 March, Poughkeepsie, NY, USA (pp. 350–354). New York: ACM.

Dascalu, M., Trausan-Matu, S., Dessus, P., & McNamara, D. S. (2015b). Dialogism: A framework for CSCL and a signature of collaboration. In O. Lindwall, P. Häkkinen, T. Koschmann, P. Tchounikine, & S. Ludvigsen (Eds.), *Proceedings of the 11th International Conference on Computer-Supported Collaborative Learning* (CSCL 2015), 7–11 June 2015, Gothenburg, Sweden (pp. 86–93). International Society of the Learning Sciences.

Dascalu, M., Trausan-Matu, S., McNamara, D. S., & Dessus, P. (2015). ReaderBench: Automated evaluation of collaboration based on cohesion and dialogism. *International Journal of Computer-Supported Collaborative Learning*, 10(4), 395–423.

Dascalu, M., McNamara, D. S., Crossley, S. A., & Trausan-Matu, S. (2016). Age of exposure: A model of word learning. *Proceedings of the 30th Conference on Artificial Intelligence* (AAAI-16), 12–17 February 2016, Phoenix, Arizona, USA (pp. 2928–2934). Palo Alto, CA: AAAI Press.

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment*, 5(1). http://files.eric.ed.gov/fulltext/EJ843855.pdf

Dong, A. (2005). The latent semantic approach to studying design team communication. *Design Studies*, 26(5), 445–461.

Duran, N. D., Hall, C., McCarthy, P. M., & McNamara, D. S. (2010). The linguistic correlates of conversational deception: Comparing natural language processing technologies. *Applied Psycholinguistics*, 31(3), 439–462.

Elouazizi, N. (2014). Point-of-view mining and cognitive presence in MOOCs: A (computational) linguistics perspective. *EMNLP 2014*, 32. http://www.aclweb.org/anthology/W14-4105

Graesser, A. C. (in press). Conversations with AutoTutor help students learn. *International Journal of Artificial Intelligence in Education.*

Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H. H., Ventura, M., Olney, A., & Louwerse, M. M. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 180–192.

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh Metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40, 223–234.

Graesser, A. C., McNamara, D. S., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through Point & Query, AutoTutor, and iSTART. *Educational Psychologist*, 40, 225–234.

Graesser, A. C., & Person, N. K. (1994). Question asking during tutoring. *American Educational Research Journal*, *31*(1), 104–137.

Jackson, G. T., Allen, L. K., & McNamara, D. S. (2016). Common Core TERA: Text Ease and Readability Assessor. In S. A. Crossley & D. S. McNamara (Eds.), *Adaptive educational technologies for literacy instruction*. New York: Taylor & Francis, Routledge.

Jackson, G. T., Guess, R. H., & McNamara, D. S. (2010). Assessing cognitively complex strategy use in an untrained domain. *Topics in Cognitive Science*, 2, 127–137.

Jarvis, S., Bestgen, Y., Crossley, S. A., Granger, S., Paquot, M., Thewissen, J., & McNamara, D. S. (2012). The comparative and combined contributions of n-grams, Coh-Metrix indices and error types in the L1 classification of learner texts. In S. Jarvis & S. A. Crossley (Eds.), *Approaching language transfer through text classification: Explorations in the detection-based approach* (pp. 154–177). Bristol, UK: Multilingual Matters.

Johnson-Glenberg, M. C. (2007). Web-based reading comprehension instruction: Three studies of 3D-readers. In D. McNamara (Ed.), *Reading comprehension strategies: Theory, interventions, and technologies* (pp. 293–324). Mahwah, NJ: Lawrence Erlbaum Publishers.

Jurafsky, D., & Martin, J. H. (2000). *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition*. Upper Saddle River, NJ: Prentice Hall.

Jurafsky, D., & Martin, J. H. (2008). *Speech and language processing: An introduction to natural language processing, computational linguistics and speech recognition*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Koller, D., Ng, A., Do, C., & Chen, Z. (2013). Retention and intention in massive open online courses: In depth. *Educause Review*, 48(3), 62–63.

Koschmann, T. (1999). Toward a dialogic theory of learning: Bakhtin's contribution to understanding learning in settings of collaboration. In C. M. Hoadley & J. Roschelle (Eds.), *Proceedings of the 1999 Conference on Computer Support for Collaborative Learning* (CSCL '99), 12–15 December 1999, Palo Alto, California (pp. 308–313). International Society of the Learning Sciences.

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly*, 49(4), 757–786.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2), 211–240.

Landauer, T. K., Laham, D., & Foltz, P. W. (2003). Automated scoring and annotation of essays with the Intelligent Essay Assessor. In M. D. Shermis & J. Burstein (Eds.), *Automated essay scoring: A cross-disciplinary perspective* (pp. 87–112). Mahwah, NJ: Lawrence Erlbaum Associates.

Landauer, T., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Landauer, T. K., Kireyev, K., & Panaccione, C. (2011). Word maturity: A new metric for word knowledge. *Scientific Studies of Reading*, 15(1), 92–108.

Louwerse, M. M., McCarthy, P. M., McNamara, D. S., & Graesser, A. C. (2004). Variation in language and cohesion across written and spoken registers. In K. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (CogSci 2004), 4–7 August 2004, Chicago, IL, USA (pp. 843–848). Mahwah, NJ: Lawrence Erlbaum Associates.

McCarthy, P. M., Briner, S. W., Rus, V., & McNamara, D. S. (2007). Textual signatures: Identifying text-types using latent semantic analysis to measure the cohesion of text structures. In A. Kao & S. Poteet (Eds.), *Natural language processing and text mining* (pp. 107–122). London: Springer-Verlag UK.

McKeown, M. G., Beck, I. L., & Blake, R. G. K. (2009). Rethinking reading comprehension instruction: A comparison of instruction for strategies and content approaches. *Reading Research Quarterly*, 44, 218–253.

McNamara, D. S. (2011). Computational methods to extract meaning from text and advance theories of human cognition. *Topics in Cognitive Science*, 2, 1–15.

McNamara, D. S., Boonthum, C., Levinstein, I. B., & Millis, K. (2007). Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In T. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *Handbook of latent semantic analysis* (pp. 227–241). Mahwah, NJ: Lawrence Erlbaum Associates.

McNamara, D. S., Crossley, S. A., & Roscoe, R. D. (2013). Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45, 499–515.

McNamara, D. S., Graesser, A. C., McCarthy, P., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge, UK: Cambridge University Press.

McNamara, D. S., & Kintsch, W. (1996). Learning from text: Effects of prior knowledge and text coherence. *Discourse Processes*, 22, 247–288.

McNamara, D. S., Levinstein, I. B., & Boonthum, C. (2004). iSTART: Interactive strategy trainer for active reading and thinking. *Behavioral Research Methods, Instruments, & Computers*, 36, 222–233.

McNamara, D. S., Ozuru, Y., Graesser, A. C., & Louwerse, M. (2006). Validating Coh-Metrix. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (CogSci 2006), 26–29 July 2006, Vancouver, British Columbia, Canada (pp. 573–578). Austin, TX: Cognitive Science Society.

McNamara, D. S., Raine, R., Roscoe, R., Crossley, S. A, Jackson, G. T., Dai, J., Cai, Z., Renner, A., Brandon, R., Weston, J., Dempsey, K., Carney, D., Sullivan, S., Kim, L., Rus, V., Floyd, R., McCarthy, P. M., & Graesser, A. C. (2012). The Writing-Pal: Natural language algorithms to support intelligent tutoring on writing strategies. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation, and resolution* (pp. 298–311). Hershey, PA: IGI Global.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representation in vector space. In *Workshop at ICLR*. Scottsdale, AZ. https://arxiv.org/abs/1301.3781

Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4), 235–244.

Moon, S., Potdar, S., & Martin, L. (2014). Identifying student leaders from MOOC discussion forums through language influence. *EMNLP 2014*, 15. http://www.aclweb.org/anthology/W14-4103

Pennebaker, J. W., Booth, R. J., & Francis, M. E. (2007). Linguistic inquiry and word count: LIWC [Computer software]. Austin, TX: liwc.net.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The development and psychometric properties of LIWC2015. UT Faculty/Researcher Works. https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf?sequence=3

Roscoe, R. D., & McNamara, D. S. (2013). Writing Pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, 105, 1010–1025.

Rudner, L. M., Garcia, V., & Welch, C. (2006). An evaluation of IntelliMetric' essay scoring system. *The Journal of Technology, Learning and Assessment*, 4(4). https://ejournals.bc.edu/ojs/index.php/jtla/article/download/1651/1493

Seaton, D. T., Bergner, Y., Chuang, I., Mitros, P., & Pritchard, D. E. (2014). Who does what in a massive open online course? *Communications of the* ACM, 57(4), 58–65.

Shermis, M. D., Burstein, J., Higgins, D., & Zechner, K. (2010). Automated essay scoring: Writing assessment and instruction. *International Encyclopedia of Education*, 4, 20–26.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge* (pp. 451–473). Cambridge, MA: MIT Press.

Teplovs, C. (2008). The knowledge space visualizer: A tool for visualizing online discourse. In G. Kanselaar, V. Jonker, P. A. Kirschner, & F. Prins (Eds.), *Proceedings of the International Society of the Learning Sciences* 2008: *Cre8 a learning world*. Utrecht, NL: International Society of the Learning Sciences. http://chris.ikit.org/ksv2.pdf

Trausan-Matu, S., Rebedea, T., Dragan, A., & Alexandru, C. (2007). Visualisation of learners' contributions in chat conversations. In J. Fong & F. L. Wang (Eds.), *Blended learning* (pp. 217–226). Singapore: Pearson/Prentice Hall.

Trausan-Matu, S., Stahl, G., & Sarmiento, J. (2007). Supporting polyphonic collaborative learning. *Indiana University Press, E-service Journal*, 6(1), 58–74.

Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59, 433–460. http://www.loebner.net/Prizef/TuringArticle.html

Valenti, S., Neri, F., & Cucchiarelli, A. (2003). An overview of current research on automated essay grading. *Journal of Information Technology Education: Research*, 2(1), 319–330.

Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Evaluative misalignment of 10th-grade student and teacher criteria for essay quality: An automated textual analysis. *Journal of Writing Research*, 5, 35–59.

Weigle, S. C. (2013). English as a second language writing and automated essay evaluation. In M. D. Shermis & J. Burstein (Eds.), *Handbook of automated essay evaluation: Current applications and new directions* (pp. 36–54). London: Routledge.

Wen, M., Yang, D., & Rosé, C. P. (2014a). Linguistic reflections of student engagement in massive open online courses. In E. Adar & P. Resnick (Eds.), *Proceedings of the 8th International* AAAI *Conference on Weblogs and Social Media* (ICWSM '14), 1–4 June 2014, Ann Arbor, Michigan, USA. Palo Alto, CA: AAAI Press. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8057/8153

Wen, M., Yang, D., & Rosé, C. P. (2014b). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 185–192). International Educational Data Mining Society.

Wilson, M. D. (1988). The MRC psycholinguistic database: Machine-readable dictionary (Version 2). *Behavioral Research Methods, Instruments, and Computers*, 201, 6–11.

Wong, B. Y. L. (1985). Self-questioning instructional research: A review. *Review of Educational Research*, 55, 227–268.

Xi, X. (2010). Automated scoring and feedback systems: Where are we and where are we heading? *Language Testing*, 27(3), 291–300.

Chapter 9: Discourse Analytics

Carolyn Penstein Rosé

Language Technologies Institute and Human–Computer Interaction Institute, Carnegie Mellon University, USA

DOI: 10.18608/hla17.009

ABSTRACT

This chapter introduces the area of discourse analytics (DA). Discourse analytics has its impact in multiple areas, including offering analytic lenses to support research, enabling formative and summative assessment, enabling of dynamic and context sensitive triggering of interventions to improve the effectiveness of learning activities, and provision of reflection tools such as reports and feedback after learning activities in support of both learning and instruction. The purpose of this chapter is to encourage both an appropriate level of hope and an appropriate level of skepticism for what is possible while also exposing the reader to the breadth of expertise needed to do meaningful work in this area. It is not the goal to impart the needed expertise. Instead, the goal is for the reader to find his or her place within this scope to discern what kinds of collaborators to seek in order to form a team that encompasses sufficient breadth. We begin with a definition of the field, casting a broad net both theoretically and methodologically, explore both representational and algorithmic dimensions, and conclude with suggestions for next steps for readers who are interested in delving deeper.

Keywords: Discourse analytics, collaborative learning, machine learning, analysis tools

Discourse analytics (DA) is one area within the field of learning analytics (LA; Buckingham Shum, 2013; Buckingham Shum, de Laat, de Liddo, Ferguson, & Whitelock, 2014). It includes processing of open response questions in educational contexts, and a large proportion of research in the area focuses on assessment of writing, but it encompasses more than that, including analysis of discussions occurring in discussion forums, chat rooms, microblogs, blogs, and even wikis. We consider LA broadly as learning about learning by listening to learners learn, with our listening normally assisted by data mining and machine learning technologies, though the published work in the area may precede but not yet include automation in all cases (Knight & Littleton, 2015; Milligan, 2015). Furthermore, we consider that what makes this area distinct is that the listening focuses on natural language data in all of the streams in which that data is produced.

This chapter offers a very brief introduction to this area situated within the field of LA broadly. DA is an area that has alternately suffered from two dangerous misconceptions. The first is an extreme over-expectation fuelled by the desire of many to have an off-the-shelf solution that will do their analysis work for them at the click of a button. Those falling prey to this misconception are almost certainly doomed to disappointment. Making effective use of either the most simple or the most powerful modelling technologies requires a lot of preparation, effort, and expertise. The second misconception is an extreme skepticism, sometimes resulting from disappointments arising from starting with the first misconception, or other times coming from a deep enough understanding of the complexities of discourse that it is difficult to get past the understanding that no computer could ever fully grasp the nuances that are there. While it is true that discourse is incredibly complex, it is still true that there are meaningful patterns that state-of-the-art modelling approaches are able to identify. Much published work from recent Learning Analytics and Knowledge and related conferences that illustrate the state-of-the-art are cited throughout this chapter. A recent survey on computational sociolinguistics tells the story from the perspective of the field of language technologies (Nguyen, Dogruöz, Rosé, & de Jong, in press), and might be of interest to dedicated readers.

The hope of this chapter is that it provides helpful pointers to readers who want to dig a little further. Two previous workshops on the topic of DA survey the foundational work within the LA community (Buckingham Shum, 2013; Buckingham Shum et al., 2014). An extensive overview of issues and methods situated more narrowly within the field of computer-supported collaborative learning can be found in three earlier published journal articles (Rosé et al., 2008; Mu, Stegman, Mayfield, Rosé, & Fischer, 2012; Gweon, Jain, McDonough, Raj, & Rosé, 2013). A short course in the area can be found in the text-mining unit of the Fall 2014 Data, Analytics, and Learning[1] MOOC offered on the edX platform. Other resources will be presented at the end of this chapter.

In this chapter, we are interested in the natural language uttered during episodes of learning. We seek to be theoretically and methodologically inclusive. Much of the existing work on discourse analytics views learning and its connection with language from a cognitive lens, in other words, seeking categories of language behaviour whose presence in a discourse makes predictions about learning gains because of the connection between the associated discourse processes and cognitive processes associated with learning. In this chapter, we seek to view learning and its connection with language through a social lens in order to leverage the important interplay between the cognitive and social factors in learning (Hmelo-Silver, Chinn, Chan, & O'Donnell, 2013; O'Donnell & King, 1999). For example, we seek to identify discourse processes that reveal underlying dispositions, attitudes, and relationships that play a supporting (or sometimes interfering) role in the learning interactions. Regardless of the situation in which it is uttered, natural language is deeply personal and deeply cultural. Embedded within it are artifacts of our personal experiences and those of generations that came before us. The details of language choices provide clues about the identities we purposefully project as well as sometimes those we seek to hide or even those of which we are not consciously aware. They project assumptions about and attitudes towards our audience and our positioning with respect to our audience, or sometimes just assumptions we want our audience to think we are making. We use these choices as currency in an economy of relationships in which we seek to achieve goals that we have adopted (Ribeiro, 2006).

With this understanding, as we use computation as a lens to aid in our listening to learners, we must acknowledge that we are always abdicating some of the responsibility for interpretation to the technologies that sit between us and the learning process, including whatever was lost or transformed in the recording into some digital form, and the further reduction and transformation that occurred during the application of the analytic technology (Morrow & Brown, 1994). With that caveat in mind, in this chapter we will focus heavily on questions of model interpretation and assessment of validity.

SCOPE AND FOCUS OF THIS CHAPTER

When one initially thinks about analytics, algorithms immediately pop to mind (Witten, Frank, & Hall, 2011). However, it is important to take a lesson from applied statistics and instead think about representation first. At the heart of DA work is a focus on representation of the data. Machine learning models cannot be applied directly to texts. Rather, the predictor features must be extracted from the text. These predictor features can be conceived of as questions: "Is __ found in the text?" or "How many times is __ found in the text?" If each feature is one of these questions, then for each instance, the feature value is the answer to the question. Interested readers can get a good feel for the breadth of simple features that can readily be extracted from text and what impact they have on predictive accuracy of classification models by experimenting with the publically available LightSIDE tool bench[2] (Mayfield & Rosé, 2013; Gianfortoni, Adamson, & Rosé, 2011), a freely available, off-the-shelf workbench with an extensive user's manual, example data sets, instructions about process, and contact information for researchers who are willing to offer help.

The key to success with modelling technologies applied to text is to ask the right questions, which produce meaningful clues. Thinking about this question begins by considering how language is structured. Though on the surface language may appear to the naked eye as a monolithic, unstructured whole, the fact is that it is composed of multiple layers of structure, each described within a separate area of linguistics. An introductory survey of a linguistics textbook (O'Grady, Archibald, Aronoff, & Rees-Miller, 2009) would be a valuable resource for researchers desiring to get into this area of LA. At the finest grain is the sound structure level, referred to as phonology and phonetics. Here the basic sound units of a language and how they fit together into the syllabic structure of a language are described. A basic alphabet of sounds comprise the set of phonemes, but within dialects these may be

[1] https://www.edx.org/course/data-analytics-learning-utarlington-link5-10x

[2] http://lightsidelabs.com/research/

pronounced in particular ways, which carry social significance because of their association with a host of socially relevant variables such as ethnicity, socioeconomic status, and region. Just above that level, the inner structure of words is described in a layer referred to as morphology. This is where systems of affixes we learn in our grammar classes come into the picture, which change the tenses on verbs or number on nouns, among other things. Above that is the level of syntax, where the grammatical structure of whole sentences is described. Also at the level of a sentence is the area of semantics, which describes how meaning is composed through fixed expressions, by convention, or by composing smaller units, guided through syntax, and referencing low level semantic units at the level of lexical semantics. Above the sentence level is the level of discourse, where we find rhetorical strategies among other aspects of structure. While these technical terms might be unfamiliar to many readers, they may provide useful search terms for readers who desire to find relevant resources for further reading.

If one traces the history of several areas in which natural language data has been the target of automated analysis, we hear the same refrain, namely the key to valid modelling is design of meaningful representations. The hope in including this example in this chapter is that readers can be spared from learning the same lesson the hard way. Taking one of the earliest cases where this lesson about DA was well learned was that of automated essay scoring (Page, 1966; Shermis & Hammer, 2012). The earliest approaches used simple models, like regression, and simple features, such as counting average sentence length, number of long words, and length of essay. These approaches were highly successful in terms of reliability of assignment of numeric scores (Shermis & Burstein, 2013); however, they were criticized for lack of validity in their usage of evidence for assessment. In later work, the focus shifted to identification of features more like what instructors included in their own rubrics for scoring writing. This investigation led to inclusion of content focused features, including techniques akin to factor analysis such as latent semantic analysis (LSA: Foltz, 1996) or latent Dirichlet allocation (LDA; Blei, Ng, & Jordan, 2003; Griffiths & Steyvers, 2004) to aid in content based assessments, though these still fall prey to problems with unigram features since they are also usually grounded in a unigram language representation. Other factor analytic language analysis approaches such as CohMetrix (McNamara & Graesser, 2012) have recently been used for assessment of student writing along multiple dimensions, including such factors as cognitive complexity. In highly causal domains that build in some level of syntactic structural analysis, CohMetrix has shown benefits (Rosé & VanLehn, 2005). In science education, success with assessment of open-ended responses has been achieved with LightSIDE (Nehm, Ha, & Mayfield, 2012; Mayfield & Rosé, 2013).

At this point, it is useful to return to the tension between the over- and under-expectation of DA. If we think about the challenges in identifying appropriate, meaningful features, we must come to terms with the limitations of the lenses we construct through modelling tools. The analytic technologies applied in DA may serve as a lens in the hands of researchers or practitioners that sits between them and the episodes of learning that occur within the world, or they may be a filter that mediates the interaction between learners and instructors, between learners, or between learners and learning technologies. Lenses are useful precisely because they do not simply transfer the exact details of the world viewed through them. Instead they accentuate aspects of those images that would not as effectively been seen without them. That is what we need them to do. At the same time, they obscure other details that are deemed less interesting by design. Lenses always distort. But in order to use them in a valid way, we must understand what each accentuates and obscures so that we can select an appropriate lens, and so we can interpret what we see in a valid way, always questioning how the picture would be different without it or with a different lens. Thus, from the beginning, we would caution those who consume the research in this area, develop these lenses, or actively apply them in research or practice, to be wary of what is inevitably lost or transformed in the process of application. Now this chapter will turn its attention to specific areas within the scope of DA.

REPRESENTATION OF TEXT

Key decisions that strongly influence how the data will appear through the analytic lens are made at the representation stage. At this stage, text is transformed from a seemingly monolithic whole to a set of features that are said to be extracted from it. Each feature extractor asks a question of the text, and the answer that the text gives is the value of the corresponding feature within the representation. Imagine that all you knew about a person was the set of answers to questions posed during a game of twenty questions, and now your task is to classify that person into a number of social categories of interest. If the questions are carefully constructed, you may be able to make an accurate prediction; nevertheless, you must acknowledge that much information and insight into that person as an individual will have been lost in the process. Once information is lost at this important stage in the process, it cannot be recovered through application of an algorithm, no matter how advanced and generally effective that algorithm is. Thus, we emphasize throughout this chapter the importance of careful decision making about representation, careful reflection about interpretation, and careful questioning of the validity of inferences made. While readers new to this area may find these caveats somewhat illusive, they will become clearer with experience.

Overview

Unigram features are the most typical feature extractors used in text mining problems. In the case of a unigram feature space, for each word appearing within the set of texts in the training data, there will be a corresponding feature that asks about the presence of that word within each text. While unigram feature spaces frequently achieve reasonably high performance, the models often fail to generalize beyond data collected under very similar circumstances to that of the training data. The reason for the lack of generalization is that these unigram models essentially memorize for each class value label in a superficial fashion what kinds of things people talk about in the set of instances associated with that label in the training data. If there is some consistency in that, then it can be learned by these models, but that consistency rarely generalizes very far. Generalization comes when the features extracted come from a relevant layer of structure.

The purpose of the feature-based representation of text is frequently to enable predictive modelling for classification or numerical assessment, where the objective is to achieve this predictive modelling with the highest possible accuracy (Rosé et al., 2008; McLaren et al., 2007; Allen, Snow, McNamera, 2015). This orientation will be the focus of this section. However, it is important to note that in some work within the broad area of DA, the representation work is the focus, and meaning is made of the identified predictive features, and thus the predictive modelling, if any, serves mainly as a validation of the meaningfulness of the identified features (Simsek, Sandor, & Buckingham Shum, 2015; Dascalu, Dessus, McNamera, 2015; Snow, Allen, Jacovina, Perret, McNamera, 2015).

With respect to predictive modelling for classification, in this vector-based comparison, the chosen features should make instances that are of different categories look far apart within the vector space, and instances that are of the same category look close within the vector space. This principle can also be used to troubleshoot a text representation. Features that either make instances that should be classified the same way look different or make instances that should be classified differently look similar are very likely to cause confusion in the classifications made by models trained using representations that include those features. The problem is often either ambiguous features (i.e., features that mean different things in different contexts, but the representation does not enable leveraging that context in order to disambiguate) or fragmentation (i.e., the same abstract feature is being represented by several more specific features, some of which are missing or too sparse in your data). It may also be that the most meaningful features are simply missing from your feature space, and other features, which may correlate with the meaningful ones within the specific data used as training data, will often "steal the weight," which ends up being counter-productive when the model is applied to new data where the spurious correlations between the meaningful features and less meaningful features may not exist or may be different.

Case Study

In order to illustrate the thinking that goes into representation of text for DA, we will start with a common example, namely analysis of affect in text, otherwise known as *sentiment analysis* (Pang & Lee, 2008). It is one of the most heavily marketed applications of text mining, and it is frequently the first thing researchers think to apply to their text data when faced with analyzing it. We will begin by introducing some issues in this area of text analytics and conclude with an investigation of what these analytics do or do not offer in terms of explaining patterns of attrition in MOOCs, where one might reasonably expect to see more expressions of negative affect from students who are struggling and ultimately drop out. We will see that the picture is far more complex than that (Wen, Yang, & Rosé, 2014a). In leading the reader through this case study, the hope is that the reader will see how one might progress through cycles of data analysis from pre-conceptions that start out overly simplistic, but become more informed through iteration. The most interesting work in the area of DA, or any area of analytics applied to rich, relatively unstructured data, will follow a similar storyline.

Simplistic treatments of sentiment identify texts as exhibiting either a positive or negative sentiment, and rely on an association between words and this affective judgment. Thus, much work has gone into the construction of sentiment lexicons, which associate words with a positivity or negativity score. The area of sentiment analysis is well developed, gaining substantial representation in industry, providing services to businesses related to marketing issues. Nevertheless, the limitations of the technology are clear. Furthermore, what is learned from examination of the linguistic literature is that much about attitude is not conveyed in text through words that are specifically positive or negative (Martin & White, 2005). This can be illustrated with the following example related to the weather. A statement such as "The weather is beautiful today" contains the required positive word; however, "The sun is shining" is only obviously positive if one knows that typically sunny days are preferred over rainy days. "It's a great day for staying indoors," indicates that the weather is not so good, despite the presence of a positive word. "My rain boots are feeling neglected," could easily be taken as a positive comment about the weather despite the presence of a negative word.

Now we will investigate situations more close to home where the approach may fall short. Because sentiment analysis is one of the most widely known and widely used language technologies by researchers and practitioners in other fields who are interested in text, it is not surprising that analysis of forum data from MOOCs is one area where we find applications of this technology, and thus that work will be a convenient case study. The rationale for its application was that discussion forum data may be useful for understanding better how, why, and when students drop out of MOOCs, with the idea that students may drop out because they are dissatisfied with a course, and that dissatisfaction should be visible using sentiment analysis as a lens. In an early such investigation, however, Ramesh, Goldwasser, Huang, Daumé, and Getoor (2013) found no relation between overall sentiment expressed by students (as assessed using a completely automated method) and their associated probability of course completion. Adamopoulos (2013) developed a sentiment related assessment method to measure sentiment associated with different course affordances in order to understand what students express their attitudes about in course discussion forums. They used a combination of automatically identified sentiment expressions paired with a grounded theory approach to identify themes in the course aspects mentioned in connection with attitudes. With this more detailed view, they were able to identify that not attitude in general, but attitude towards the professor, the assignments, and other course materials had the strongest association with dropout. In more recent work (Wen et al., 2014a), we pushed the automated analysis further, increasing the accuracy of sentiment measurement, and contrasting sentiment expressed by a student versus sentiment they were exposed to as well as contrasting sentiment at the student level with sentiment at the course level. In this work, the exact connection between sentiment-related variables and dropout depended upon the nature of the course.

With more probing, it became clear that a far more nuanced way of characterizing affect in posts was needed. For example, negative affect expressed in purely social exchanges might be disclosure, leading to enhanced emotional connection. Problem talk in a problem-solving course might just indicate engagement with the material. Negative affect words, expressions, and images may come up in a literature course where stories about unfortunate or stressful events are discussed, and yet that expressed sentiment might have nothing to do with a student's feeling about the experience of reading that material or even discussing that material. We conclude that sentiment analysis is not as simple as counting positive and negative words. Individual words are not enough evidence of attitude, context matters. Some rhetorical strategies combine negative and positive comments in the same review, and sometimes sentiment is expressed indirectly. Nuances like this observed through qualitative analysis must be taken into account when representing your data.

UNSUPERVISED METHODS

A variety of factor analytic (Garson, 2013; Loehlin, 2004) and latent variable analysis techniques (Skrondal & Rabe-Hesketh, 2004; Collins & Lanza, 2010) have been popular in the area. These may be unsupervised (i.e., not requiring pre-assigned labels), supervised (i.e., requiring examples to have pre-defined labels), or lightly supervised (i.e., requiring some external guidance to learning algorithms, but not requiring a pre-assigned label for every example). In this section, we focus on unsupervised methods. The most popular such techniques in the education space include factor analytics approaches like latent semantic analysis (LSA: Foltz, 1996) or structured latent variable models like latent Dirichlet allocation or LDA (Blei et al., 2003) mentioned briefly above. Thus, here we delve slightly deeper into the details and discuss strengths and limitations. In recent work in LA, unsupervised approaches have been used for exploratory data analysis (Joksimović et al., 2015; Sekiya, Marsuda, & Yamaguchi, 2015; Chen, Chen, & Xing, 2015), sometimes paired with visualization techniques (Hsiao & Awasthi, 2015), or alternating with or building on hand analysis (Molenaar & Chiu, 2015; Ezen-Can, Boyer, Kellog, & Booth, 2015). These modelling technologies have widely been used because researchers think of them as approximating an analysis of textual meaning. The reality is that they are much less apt at doing so than the prevailing view would have one believe. These tools do indeed have their place in the arsenal of DA tools. However, the hope of this chapter is to raise the curiosity of the reader to dig a little deeper in order to foster an appropriate scepticism, as described above.

Topic modelling approaches have become very popular for modelling a variety of characteristics of unlabelled data. A well-known and widely used approach is LDA (Blei et al., 2003), which is a generative model effective for uncovering the thematic structure of a document collection. Hidden Markov modelling (HMM) and other sequence modelling approaches are becoming popular for capturing progressions in student experiences (Molenaar & Chiu, 2015). Sometimes these approaches are combined in order to identify how language expression changes in predictable ways over time in terms of the representations of thematic content (Jo & Rosé, 2015). Statistical approaches such as these are meant to capture regularities. They are most valuable as tools in methodologies that value data reduction and simplification. Because they dismiss as noise the unusual occurrences within the data, they are less valuable in methodologies that seek unusual happenings that challenge assumptions. Though one might adopt an anomaly detection approach to identify instances that violate assumptions as a way of identifying such examples, in practice the examples found are more likely to be unusual in ways that are not necessarily interesting from the standpoint of challenging assumptions of theoretical import.

LDA works by associating words together within a latent word class that frequently occur together within the same document. The learned structure is more complex than traditional latent class models, where the latent structure is a probabilistic assignment of each whole data point (which is a document) to a single latent class (Collins & Lanza, 2010). An additional layer of structure is included in an LDA model such that words within documents are probabilistically assigned to latent classes in such a way that data points can be viewed as mixtures of latent classes. This structure is important for topic analysis. By allowing the representation of documents as arbitrary mixtures of latent word classes, it is possible then to keep the number of latent classes down to a manageable size while still capturing the flexible way themes can be blended within individual documents. Each latent word class is represented as a distribution of words. The words that rank most highly in the distribution are those treated as most characteristic of the associated latent class, or topic.

Because LDA is an unsupervised language processing technique, it would not be reasonable to expect that the identified themes would exactly match human intuition about organization of topic themes, and yet as a technique that models word co-occurrence associations, it can be expected to identify some things that would be expected to be associated. At heart, LDA is a data reduction technique. Its strengths lie in identification of word associations that are very common in a corpus, which frequently correspond to common themes. However, the common themes do not necessarily have a one-to-one correspondence with the themes of interest. Unfortunately, that means within the resulting representation, there will not be a distinct representation for those themes of interest that are not common. Similarly, unusual phrasings of common ideas will also typically fail to map to an intuitive representation within the LDA space. Representation of the textual data is also an important consideration. Typically, LDA models are computed over feature spaces composed of individual word features. Thus, whatever is not captured by individual words will not be accessible to the model.

SUPERVISED METHODS

At the other end of the spectrum are supervised methods. Taking a somewhat overly simplistic view, supervised machine learning methods are typically algorithms that operate over sets of vectors that associate a collection of predictor features, often referred to as attributes, with an outcome feature, often referred to as a class value. Recently, applications of supervised machine learning have been applied to the problem of assessment of learning processes in discussion. This problem is referred to as automatic collaborative-learning process analysis. Automatic analysis of collaborative processes has value for real-time assessment during collaborative learning, for dynamically triggering supportive interventions in the midst of collaborative-learning sessions, and for facilitating efficient analysis of collaborative-learning processes at a grand scale. This dynamic approach has been demonstrated to be more effective than an otherwise equivalent static approach to support (Kumar, Rosé, Wang, Joshi, & Robinson, 2007). Early work in automated collaborative learning process analysis focused on text-based interactions and click stream data (Soller & Lesgold, 2007; Erkens & Janssen, 2008; Rosé et al., 2008; McLaren et al., 2007; Mu et al., 2012). Early work towards analysis of collaborative processes from speech has begun to emerge as well (Gweon et al., 2013; Gweon, Agarwal, Udani, Raj, & Rosé, 2011). A consistent finding is that representations motivated by theoretical frameworks from linguistics and psychology show particular promise (Rosé & Tovares, in press; Wen, Yang, & Rosé, 2014b; Gweon et al., 2013; Rosé & VanLehn, 2005). We have already mentioned the LightSIDE tool bench as a good place to start getting experience in this area.

MOVING AHEAD

Readers who are interested in getting more familiar with the area of DA would benefit from digging first into some foundational literature. It is grounded in the fields of linguistics (Levinson, 1983; O'Grady et al., 2009), discourse analysis (Martin & Rose, 2003; Martin & White, 2005; Biber & Conrad, 2011), and language technologies (Manning & Schuetze, 1999; Jurafsky & Martin, 2009; Jackson & Moulinier, 2007).

REFERENCES

Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. *Proceedings of the 34th International Conference on Information Systems: Reshaping Society through Information Systems Design* (ICIS 2013), 15–18 December 2013, Milan, Italy. http://aisel.aisnet.org/icis2013/proceedings/BreakthroughIdeas/13/

Allen, L., Snow, E., & McNamera, D. (2015). Are you reading my mind? Modeling students' reading comprehension skills with natural language processing techniques. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 246–254). New York: ACM.

Biber, D., & Conrad, S. (2011). *Register, Genre, and Style*. Cambridge, UK: Cambridge University Press.

Blei, D., Ng, A., & Jordan, M. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.

Buckingham Shum, S. (2013). *Proceedings of the 1st International Workshop on Discourse-Centric Learning Analytics* (DCLA13), 8 April 2013, Leuven, Belgium.

Buckingham Shum, S., de Laat, M., de Liddo, A., Ferguson, R., & Whitelock, D. (2014). *Proceedings of the 2nd International Workshop on Discourse-Centric Learning Analytics* (DCLA14), 24 March 2014, Indianapolis, IN, USA.

Chen, B., Chen, X., & Xing, W. (2015). "Twitter archaeology" of Learning Analytics and Knowledge conferences. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 340–349). New York: ACM.

Collins, L., & Lanza, S. T. (2010). *Latent class and latent transition analysis with applications in the social, behavioral, and health sciences*. Wiley.

Dascalu, M., Dessus, P., & McNamera, D. (2015). Discourse cohesion: A signature of collaboration. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 350–354). New York: ACM.

Erkens, G., & Janssen, J. (2008). Automatic coding of dialogue acts in collaboration protocols. *International Journal of Computer-Supported Collaborative Learning*, 3, 447–470.

Ezen-Can, A., Boyer, K., Kellog, S., & Booth, S. (2015). Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 146–150). New York: ACM.

Foltz, P. (1996). Latent semantic analysis for text-based research. *Behavior Research Methods, Instruments, & Computers*, 28(2), 197–202.

Garson, G. D. (2013). Factor Analysis. Asheboro, NC: Statistical Associates Publishing. http://www.statisticalassociates.com/factoranalysis.htm

Gianfortoni, P., Adamson, D., & Rosé, C. P. (2011). Modeling stylistic variation in social media with stretchy patterns. *Proceedings of the 1st Workshop on Algorithms and Resources for Modeling of Dialects and Language Varieties* (DIALECTS '11), 31 July 2011, Edinburgh, Scotland (pp. 49–59). Association for Computational Linguistics.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*, 101, 5228–5235.

Gweon, G., Jain, M., McDonough, J., Raj, B., & Rosé, C. P. (2013). Measuring prevalence of other-oriented transactive contributions using an automated measure of speech style accommodation. *International Journal of Computer Supported Collaborative Learning*, 8(2), 245–265.

Gweon, G., Agarwal, P., Udani, M., Raj, B., & Rosé, C. P. (2011). The automatic assessment of knowledge integration processes in project teams. *Proceedings of the 9th International Conference on Computer-Supported Collaborative Learning, Volume 1: Long Papers* (CSCL 2011), 4–8 July 2011, Hong Kong, China (pp. 462–469). International Society of the Learning Sciences.

Hmelo-Silver, C., Chinn, C., Chan, C., & O'Donnell, A. (2013). *The International Handbook of Collaborative Learning*. Routledge.

Hsiao, I., & Awasthi, P. (2015). Topic facet modeling: Semantic and visual analytics for online discussion forums. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 231–235). New York: ACM.

Jackson, P., & Moulinier, I. (2007). Natural language processing for online applications: Text retrieval, extraction, and categorization. Amsterdam: John Benjamins Publishing Company.

Jo, Y., Loghmanpour, N., & Rosé, C. P. (2015). Time series analysis of nursing notes for mortality prediction via state transition topic models. *Proceedings of the 24th ACM International Conference on Information and Knowledge Management* (CIKM '15), 19–23 October 2015, Melbourne, VIC, Australia (pp. 1171–1180). New York: ACM.

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015). What do cMOOC participants talk about in social media? A topic analysis of discourse in a cMOOC. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 156–165). New York: ACM.

Jurafsky, D., & Martin, J. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. Pearson.

Knight, S., & Littleton, K. (2015). Developing a multiple-document-processing performance assessment for epistemic literacy. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 241–245). New York: ACM.

Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., & Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Proceedings of the 13th International Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts that Work* (AIED 2007), 9–13 July 2007, Los Angeles, CA, USA (pp. 383–390). IOS Press.

Levinson, S. (1983). Conversational structure. *Pragmatics* (pp. 284–286). Cambridge, UK: Cambridge University Press.

Loehlin, J. C. (2004). Latent variable models: An introduction to factor, path, and structural equation analysis. Routledge.

Manning, C., & Schuetze, H. (1999). *Foundations of statistical natural language processing*. MIT Press.

Martin, J., & Rose, D. (2003). Working with discourse: Meaning beyond the clause. Continuum.

Martin, J., & White, P. (2005). *The language of evaluation: Appraisal in English*. Palgrave.

Mayfield, E., & Rosé, C. P. (2013). LightSIDE: Open source machine learning for text accessible to non-experts. In M. D. Shermis & J. Burstein (Eds.), *Handbook of Automated Essay Grading: Current Applications and New Directions* (pp. 124–135). Routledge.

McLaren, B., Scheuer, O., De Laat, M., Hever, R., de Groot, R., & Rosé, C. P. (2007). Using machine learning techniques to analyze and support mediation of student E-discussions. *Proceedings of the 13th International Conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work* (AIED 2007), 9–13 July 2007, Los Angeles, CA, USA (pp. 331–338). IOS Press.

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum (Eds.), *Applied Natural Language Processing: Identification, Investigation, and Resolution* (pp. 188–205). Hershey, PA: IGI Global.

Milligan, S. (2015). Crowd-sourced learning in MOOCs: Learning analytics meets measurement theory. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 151–155). New York: ACM.

Molenaar, I., & Chiu, M. (2015). Effects of sequences of socially regulated learning on group performance. *Proceedings of the 5th Internation Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 236–240). New York: ACM.

Morrow, R. A., & Brown, D. D. (1994). Deconstructing the conventional discourse of methodology: Quantitative versus qualitative methods. In R. A. Morrow & D. D. Brown (Eds.), *Critical theory and methodology: Contemporary social theory*, Vol. 3 (pp. 199–225). Thousand Oaks, CA: Sage.

Mu, J., Stegmann, K., Mayfield, E., Rosé, C. P., & Fischer, F. (2012). The ACODEA framework: Developing segmentation and classification schemes for fully automatic analysis of online discussions. *International Journal of Computer-Supported Collaborative Learning*, 138, 285–305.

Nehm, R., Ha, M., & Mayfeld, E. (2012). Transforming biology assessment with machine learning: Automated scoring of written evolutionary explanations. *Journal of Science Education and Technology*, 21, 183–196.

Nguyen, D., Dogruöz, A. S., Rosé, C. P., & de Jong, F. (in press). Computational sociolinguistics: A survey. *Computational Linguistics*.

O'Donnell, A., & King, A. (1999). *Cognitive perspectives on peer learning*. Routledge.

O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2009). *Contemporary linguistics: An introduction*. Boston/New York: Bedford/St. Martins.

Page, E. B. (1966). The imminence of grading essays by computer. *Phi Delta Kappan*, 48, 238–243.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135.

Ramesh, A., Goldwasser, D., Huang, B., Daumé III, H., & Getoor, L. (2013). Modeling learner engagement in MOOCs using probabilistic soft logic. *NIPS Workshop on Data Driven Education: Advances in Neural Information Processing Systems* (NIPS-DDE 2013), 9 December 2013, Lake Tahoe, NV, USA. https://www.umiacs.umd.edu/~hal/docs/daume13engagementmooc.pdf

Ribeiro, B. T. (2006). Footing, positioning, voice: Are we talking about the same things? In A. De Fina, D. Schiffrin, & M. Bamberg (Eds.), *Discourse and identity* (pp. 48–82). New York: Cambridge University Press.

Rosé, C. P., & Tovares, A. (2015). What sociolinguistics and machine learning have to say to one another about interaction analysis. In L. Resnick, C. Asterhan, & S. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue*. Washington, DC: American Educational Research Association.

Rosé, C. P., & VanLehn, K. (2005). An evaluation of a hybrid language understanding approach for robust selection of tutoring goals. *International Journal of Artificial Intelligence in Education*, 15, 325–355.

Rosé, C. P., Wang, Y. C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F., (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. International Journal of Computer Supported Collaborative Learning, 3(3), 237–271.

Sekiya, T., Marsuda, Y., & Yamaguchi, K. (2015). Curriculum analysis of CS departments based on CS2013 by simplified, supervised LDA. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 330–339). New York: ACM.

Shermis, M. D., & Burstein, J. (2013). *Handbook of Automated Essay Evaluation: Current Applications and New Directions*. New York: Routledge.

Shermis, M., & Hammer, B. (2012). Contrasting state-of-the-art automated scoring of essays: Analysis. Annual National Council on Measurement in Education Meeting, 14–16.

Simsek, D., Sandor, A., & Buckingham Shum, S. (2015). Correlations between automated rhetorical analysis and tutor's grades on student essays. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 355–359). New York: ACM.

Skrondal, A., & Rabe-Hesketh, S. (2004). *Generalized latent variable modeling: Multi-level, longitudinal, and structural equation models*. Chapman & Hall/CRC.

Snow, E., Allen, L., Jacovina, M., Perret, C., & McNamera, D. (2015). You've got style: Writing flexibility across time. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 194–202). New York: ACM.

Soller, A., & Lesgold, A. (2007). Modeling the process of collaborative learning. In H. U. Hoppe, H. Ogata, & A. Soller (Eds.), *The role of technology in CSCL: Studies in technology enhanced collaborative learning* (pp 63–86). Springer. doi:10.1007/978-0-387-71136-2_5

Wen, M., Yang, D., & Rosé, C. P. (2014a). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK. International Educational Data Mining Society. https://www.researchgate.net/publication/264080975_Sentiment_analysis_in_MOOC_discussion_forums_What_does_it_tell_us

Wen, M., Yang, D., & Rosé, C. P. (2014b). Linguistic reflections of student engagement in massive open online courses. *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media* (ICWSM '14), 1–4 June 2014, Ann Arbor, Michigan, USA. Palo Alto, CA: AAAI Press. http://www.cs.cmu.edu/~mwen/papers/icwsm2014-camera-ready.pdf

Witten, I. H., Frank, E., & Hall, M. (2011). *Data mining: Practical machine learning tools and techniques*, 3rd ed. San Francisco, CA: Elsevier.

Chapter 10: Emotional Learning Analytics

Sidney K. D'Mello

Departments of Psychology and Computer Science & Engineering, University of Notre Dame, USA

DOI: 10.18608/hla17.010

ABSTRACT

This chapter discusses the ubiquity and importance of emotion to learning. It argues that substantial progress can be made by coupling the discovery-oriented, data-driven, analytic methods of learning analytics (LA) and educational data mining (EDM) with theoretical advances and methodologies from the affective and learning sciences. Core, emerging, and future themes of research at the intersection of these areas are discussed.

Keywords: Affect, affective science, affective computing, educational data mining

At the "recommendation" of a reviewer of one of my papers (D'Mello, 2016), I recently sought to learn a new (for me) statistical method called generalized additive mixed models (GAMMs; McKeown & Sneddon, 2014). GAMMs aim to model a response variable with an additive combination of parametric and nonparametric smooth functions of predictor variables, while addressing autocorrelations among residuals (for time series data). At first, I was mildly *displeased* at the thought of having to do more work on this paper. *Anxiety* resulted from the thought that I might not have the time to learn and implement a new method to meet the revision deadline. So I did nothing. The *anxiety* transitioned into mild *panic* as the deadline approached. I finally decided to look into GAMMs by downloading a recommended paper. The paper had some eye-catching graphics, which piqued my *curiosity* and motivated me to explore further. The *curiosity* quickly turned into *interest* as I read more about the method, and eventually into *excitement* when I realized the power of the approach. This motivated me to slog through the technical details, which led to some intense emotions - *confusion* and *frustration* when things did not make sense, *despair* when I almost gave up, *hope* when I thought I was making progress, and eventually, *delight* and *happiness* when I actually did make progress. I then attempted to apply the method to my data by modifying some R-syntax. More *confusion*, *frustration*, and *despair* interspersed with *hope*, *delight*, and *happiness*. I eventually got it all working and wrote up the results. Some more emotions occurred during the writing and revision cycles. Finally, I was done. I felt *contentment*, *relief*, and a bit of *pride*.

As this example illustrates, there is an undercurrent of emotion throughout the learning process. This is not unique to learning as all "cognition" is tinged with "emotion". The emotions may not always be consciously experienced (Ohman & Soares, 1994), but they exist and influence cognition nonetheless. Also, emotions do not occur in a vacuum; they are deeply intertwined within the social fabric of learning. It does not take much to imagine the range of emotions experienced by the typical student whose principle occupation is learning. Pekrun and Stephens (2011) call these "academic emotions" and group them into four categories. *Achievement* emotions (contentment, anxiety, and frustration) are linked to learning activities (homework, taking a test) and outcomes (success, failure). *Topic* emotions are aligned with the learning content (empathy for a protagonist while reading classic literature). *Social* emotions such as pride, shame, and jealousy occur because education is situated in social contexts. Finally, *epistemic* emotions arise from cognitive processing, such as surprise when novelty is encountered or confusion in the face of an impasse.

Emotions are not merely incidental; they are functional or they would not have evolved (Darwin, 1872; Tracy, 2014). Emotions perform *signalling functions* (Schwarz, 2012) by highlighting problems with knowledge (confusion), problems with stimulation (boredom), concerns with impending performance (anxiety), and challenges that cannot be easily surpassed (frustration). They perform *evaluative functions* by serving as the currency by which people appraise an event in terms of its value, goal relevance, and goal congruence (Izard, 2010). Emotions perform *modulation functions* by constraining or expanding cognitive focus with negative emotions engendering narrow, bottom-up, and focused modes of processing (constrained focus) (Barth & Funke, 2010; Schwarz, 2012) in comparison to positive emotions, which facilitate broader, top-down, generative processing (expanded focus) (Fredrickson & Branigan, 2005; Isen, 2008). Indeed, emotions pervade thought as is evident by their effects on memory, problem solving, decision making, and other facets of cognition (see Clore & Huntsinger, 2007, for a review).

So what exactly is an "emotion"? Truth be told, we do not really know or at least we do not fully agree (Izard, 2010). This can be readily inferred from the most recent debates on the psychological underpinnings of emotion - a debate sometimes referred to as the "100 year old emotion war" (Lench, Bench, & Flores, 2013; Lindquist, Siegel, Quigley, & Barrett, 2013). Fortunately, there is general agreement on the following key points. Emotions are conceptual entities that arise from brain–body–environment interactions. But you will not find them by looking in the brain, the body, or the environment. Instead, emotions *emerge* (Lewis, 2005) when organism–environment interactions trigger changes across multiple time scales and at multiple levels - neurobiological, physiological, behaviourally expressive, action-oriented, and cognitive/metacognitive/subjective. The "emotion" is reflected in these changes in a manner modulated by the ongoing situational context. The same emotional category (e.g., anxiety) will manifest differently based on a triggering event (Tracy, 2014), the specific biological/cognitive/metacognitive processes involved (Gross, 2008; Moors, 2014), and sociocultural influences (Mesquita & Boiger, 2014; Parkinson, Fischer, & Manstead, 2004). For example, an anxiety-inducing event will trigger distinct "episodes" of anxiety depending on the specific circumstance (public speaking, test taking), the temporal context (one day vs. one minute before the speech), the neurobiological system (baseline arousal), and the social context (speaking in front of colleagues vs. strangers). This level of variability and ambiguity is expected because humans and their emotions are dynamic and adaptive. Rigid emotions have little evolutionary value.

Where do learning analytics (LA) and educational data mining (EDM) fit in? On one hand, given the central role of emotions in learning, attempts to analyze (or data mine) learning without considering emotion will be incomplete. On the other hand, giving the ambiguity and complexity of emotional phenomena, attempts to study emotions during learning without the methods of LA and EDM will only yield shallow insights. Fortunately, there is a body of work that adopts a data-driven analytic approach to study the incidence and influence of emotions on the processes and products of learning. In this chapter, I highlight some of the core, emerging, and future themes in this interdisciplinary research area.

Let us begin with a note on terminology. Emotion is related, but not equivalent to motivation, attitudes, preferences, physiology, arousal, and a host of other constructs often used to refer to it. Emotions are also distinct from moods and affective traits (Rosenberg, 1998). Emotion is not the same as a feeling. Hunger is a feeling but is not an emotion. Neither is pain. There is also some contention as to what constitutes an emotion. Anger is certainly an emotion, but what about confusion? Confusion has affective components (feelings of being confused, characteristic facial expressions; D'Mello & Graesser, 2014b), but there is some debate as to whether it is an emotion (Hess, 2003; Rozin & Cohen, 2003). Thus, in the remainder of this chapter, I use the more inclusive term "affective state" rather than the more restrictive term "emotion".

CORE THEMES

I selected the following four themes to highlight the use of LA/EDM methods to study affect during learning. I also review one or two exemplary studies within each theme in some depth rather than cursorily reviewing many studies. This means that many excellent studies go unmentioned, but I leave it to the reader to explore the body of work within each theme. I recommend review papers, when available, to facilitate this process.

Affect Analysis from Click-Stream Data

One of the most basic uses of LA/EDM techniques is to use the rich stream of data generated during interactions with learning technologies in order to understand learners' cognitive processes (Corbett & Anderson, 1995; Sinha, Jermann, Li, & Dillenbourg, 2014). A complementary set of insights can be gleaned when affect is included in the mix, as illustrated in the study below.

Bosch and D'Mello (in press) conducted a lab study on the affective experience of students during their first programming session. Novice students (N = 99) were asked to learn the fundamentals of computer programming in the Python language using a self-paced computerized learning environment involving a 25-minute scaffolded learning phase and a 10-minute non-scaffolded fadeout phase. All instructional activities (coding, reading text, testing code, receiving errors, etc.) were logged and videos of students' faces and computer screens were recorded. Students provided affect judgments at approximately 100 points (every 15 seconds) over the course of viewing these videos immediately after the learning session via a retrospective affect judgment protocol (Porayska-Pomsta, Mavrikis, D'Mello, Conati, & Baker, 2013). The affective states of interest were anger, anxiety, boredom, confusion, curiosity, disgust, fear, frustration, flow/engagement, happiness, sadness, and surprise. Only engagement, confusion, frustration, boredom, and curiosity occurred with sufficient frequency to warrant further analysis.

The authors examined how interaction events give rise to affective states, and how affective states trigger various behaviours. They constructed time series that interspersed interaction events (from clickstream data) and affective states (self-reports) for each student during the scaffolded learning phase. Time series modelling techniques (D'Mello, Taylor, & Graesser, 2007) were used to identify significant transitions between affective states and interaction events. The resultant model is shown as a directed graph in Figure 10.1. There were some transitions between interaction events that did not include an affective state (dashed lines). This was due to the infrequency of affect sampling (every 15 seconds) relative to other interaction events (as frequent as 1 second).

The more interesting transitions include affective states. In particular, confusion and frustration were both preceded by an incorrect solution submission (SubmitError); these affective states were then followed by a hint request (ShowHint), or by constructing code (Coding), which itself triggered confusion and frustration. Reading instructional texts (including problem descriptions) was a precursor of engagement, curiosity, boredom, and confusion but not frustration. In other words, all the key affective states were related to knowledge assimilation (reading) and construction (coding) activities. However, only confusion and frustration accompanied failure (Submit Error) and subsequent help-seeking behaviours (ShowHint), which are presumably learning opportunities. Taken together, the transition model emphasizes the key role of impasses and the resultant negative activating states of confusion and frustration to learning (D'Mello & Graesser, 2012b; VanLehn, Siler, Murray, Yamauchi, & Baggett, 2003). It also illustrates how affect is interspersed throughout the learning process.

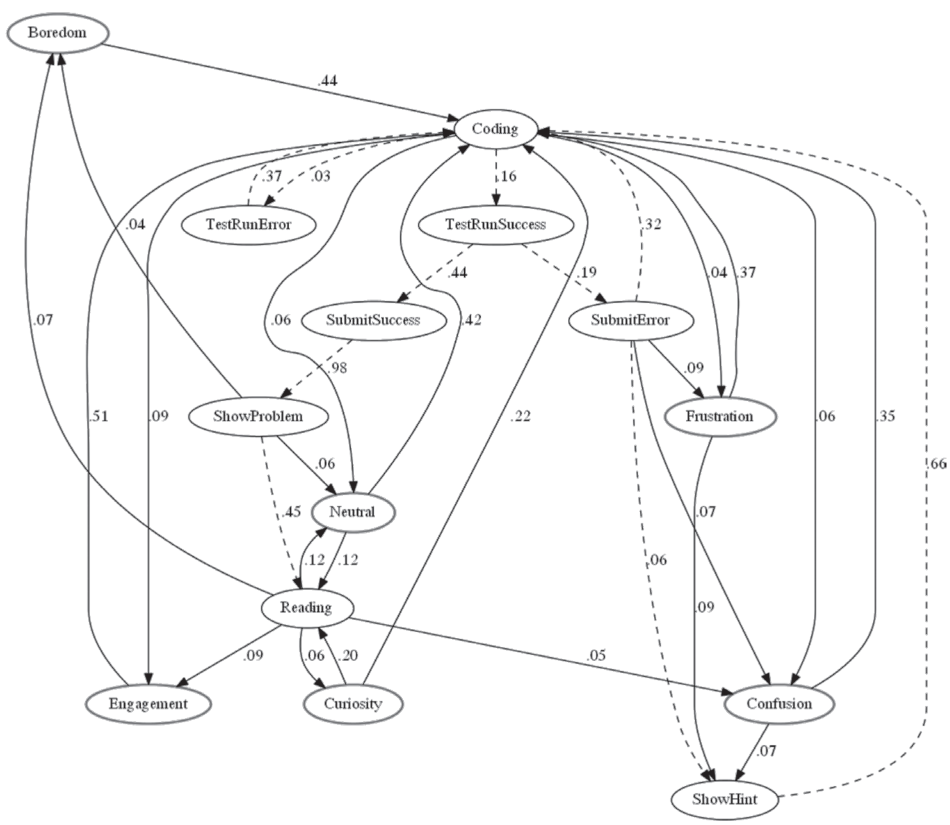


Figure 10.1. Significant transitions between affective states and interaction events during scaffolded learning of computer programming. Solid lines indicate transitions including affect. Dashed lines indicate transitions not involving affective states. ShowProblem: starting a new exercise; Reading: viewing the instructions and/or problem statement; Coding: editing or viewing the current code; ShowHint: viewing a hint; TestRunError: code was run and encountered a syntax or runtime error; TestRunSuccess: code run without syntax or runtime errors (but was not checked for correctness); SubmitError: code submitted and produced an error or incorrect answer; SubmitSuccess: code submitted and was correct.

Affect Detection from Interaction Patterns

Affective states cannot be directly measured because they are conceptual entities (constructs). However, they emerge from environment–person interactions (context) and influence action by modulating cognition. Therefore, it should be possible to "infer" affect by analyzing the unfolding context and learner actions. This line of work, referred to as "interaction-based", "log-file based", or "sensor-free" affect detection was started more than a decade ago (Ai et al., 2006; D'Mello, Craig, Sullins, & Graesser, 2006) and was recently reviewed by Baker and Ocumpaugh (2015).

As an example, consider Pardos, Baker, San Pedro, and Gowda (2013), who developed affect detectors for ASSISTments, an intelligent tutoring system (ITS) for middle- and high-school mathematics, used by approximately 50,000 students in the US as part of their regular mathematics instruction (Razzaq et al., 2005). The authors adopted a supervised learning approach to build automated affect detectors. They collected training data from 229 students while they used ASSISTments in school computer labs. Human observers provided online observations (annotations) of affect as students interacted with ASSISTments using the Baker-Rodrigo Observation Method Protocol (BROMP) (Ocumpaugh, Baker, & Rodrigo, 2012). According to this protocol, trained observers provide live annotations of affect based on observable behaviour, including explicit actions towards the interface, interactions with peers and teachers, body movements, gestures, and facial expressions. The observers coded four affective states (boredom, frustration, engaged concentration, and confusion) and two behaviours (going off-task and gaming the system). Supervised learning techniques were used to discriminate each affective state from other states (e.g., bored vs. others) using features extracted from ASSISTments log files (performance on problems, hint requests, response times, etc.). Affect detection accuracies ranged from .632 to .678 (measured with the A-prime metric [similar to area under the receiver operating characteristic curve – AUC or AUROC]) for affect and .802 to .819 for behaviours. The classifier was validated in a manner that ensured generalizability to new students from the same population by enforcing strict independence among training and testing data.

Pardos et al. (2013) also provided preliminary evidence on the predictive validity of their detectors. This was done by applying the detectors on log files from a different set of 1,393 students who interacted with ASSISTments during the 2004–2006 school years - several years before the measure was developed. Automatically measured affect and behaviour moderately correlated with standardized test scores. Further, San Pedro, Baker, Bowers, and Heffernan (2013) attempted to predict college enrollment based on the automatic detectors. They applied the detectors to existing log files from 3,707 students who interacted with ASSISTments from 2004 to 2009. College enrollment information for these students was obtained from the National Student Clearinghouse. Automatically measured affective states were significant predictors of college enrollment several years later, which is a rather impressive finding.

Affect Detection from Bodily Signals

Affect is an embodied phenomenon in that it activates bodily response systems for action. This should make it possible to infer learner affect (a latent variable) from machine-readable bodily signals (observables). There is a rich body of work on the use of bodily signals to detect affect as discussed in a number of reviews (Calvo & D'Mello, 2010; D'Mello & Kory, 2015; Zeng, Pantic, Roisman, & Huang, 2009). The research has historically focused on interactions in controlled environments, but researchers have begun to take this work into the real world, notably computer-enabled classrooms. The study reviewed below reflects one such effort by our research group and collaborators, but the reader is directed to Arroyo et al. (2009) for their pioneering work on affect detection in computer-enabled classrooms.

Bosch, D'Mello, Baker, Ocumpaugh, and Shute (2016) studied automated detection of affect from facial features in a noisy real-world setting of a computer-enabled classroom. In this study, 137 middle and high school students played a conceptual physics educational game called Physics Playground (Shute, Ventura, & Kim, 2013) in small groups for 1.5 to 2 hours across two days as part of their regular physics/physical science classes. Trained observers performed live annotations of boredom, confusion, frustration, engaged-concentration, and delight using the BROMP field observation protocol as in the ASSISTments study discussed above (Pardos et al., 2013). The observers also noted when students went off-task.

Videos of students' faces and upper bodies were recorded during game-play and synchronized with the affect annotations. The videos were processed using the FACET computer-vision program (Emotient, 2014), which provides estimates of the likelihood of 19 facial action units (Ekman & Friesen, 1978) (e.g., raised brow, tightened lips), head pose (orientation), and position (see Figure 10.2 for screenshot). Body movement was also estimated from the videos using motion filtering algorithms (Kory, D'Mello, & Olney, 2015) (see Figure 10.3). Supervised learning methods were used to build detectors of each affective state (e.g., bored vs. other states) using both facial expressions and bodily move-

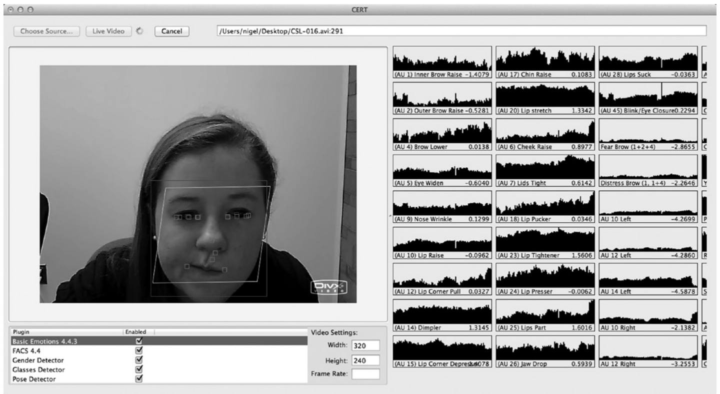


Figure 10.2. Automatic tracking of facial features using the Computer Expression Recognition Toolbox. The graphs on the right show likelihoods of activation of various facial features (e.g., brow lowered, eyelids tightening).

Figure 10.3. Automatic tracking of body movement from video using motion silhouettes. The image on the right displays the areas of movement from the video playing on the left. The graph on the bottom shows the amount of movement over time.

ments. The detectors were moderately successful with accuracies (quantified with the AUC metric as noted above) ranging from .610 to .867 for affect and .816 for off-task behaviours. Follow-up analyses confirmed that the affect detectors generalized across students, multiple days, class periods, and across different genders and ethnicities (as perceived by humans).

One limitation of the face-based affect detectors is that they are only applicable when the face can be automatically detected in the video stream. This is not always the case due to excessive movement, occlusion, poor lighting, and other factors. In fact, the face-based affect detectors were only applicable to 65% of the cases. To address this, Bosch, Chen, Baker, Shute, and D'Mello (2015) used multimodal fusion techniques to combine interaction-based (similar to previous section) and face-based detection. The interaction-based detectors were less accurate than the face-based detectors (Kai et al., 2015), but were applicable to almost all of the cases. By combining the two, the applicability of detectors increased to 98% of the cases, with only a small reduction (<5% difference) in accuracy compared to face-based detection.

Integrating Affect Models in Affect-Aware Learning Technologies

The interaction- and bodily-based affect detectors discussed above are tangible artifacts that can be instrumented to provide real-time assessments of student affect during interactions with a learning technology. This affords the exciting possibility of closing the loop by dynamically responding to the sensed affect. The aim of such affect-aware learning technologies is to expand the bandwidth of adaptivity of current learning technologies by responding

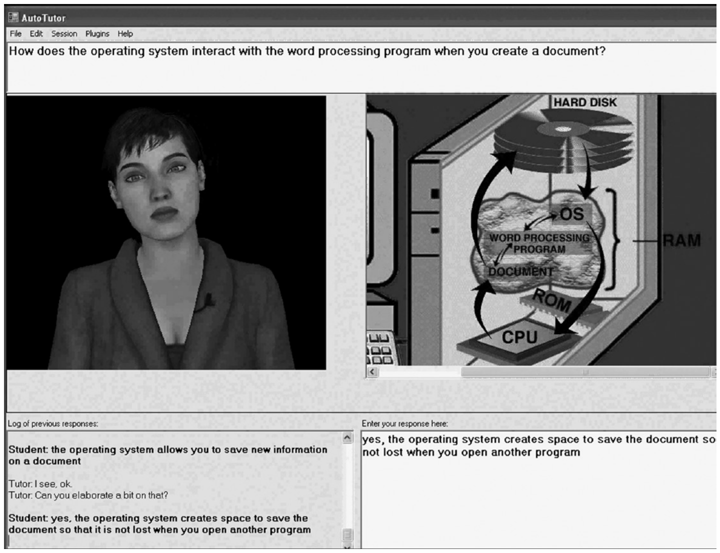


Figure 10.4. Affective AutoTutor: an intelligent tutoring system (ITS) with conversational dialogs that automatically detects and responds to learners' boredom, confusion, and frustration.

to what students feel in addition to what they think and do (see D'Mello, Blanchard, Baker, Ocumpaugh, & Brawner, 2014, for a review). Here, I highlight two such systems, the Affective AutoTutor (D'Mello & Graesser, 2012a) and UNC-ITSPOKE (Forbes-Riley & Litman, 2011).

Affective AutoTutor (see Figure 10.4) is a modified version of AutoTutor - a conversational ITS that helps students develop mastery on difficult topics in Newtonian physics, computer literacy, and scientific reasoning by holding a mixed-initiative dialog in natural language (Graesser, Chipman, Haynes, & Olney, 2005). The original AutoTutor system has a set of fuzzy production rules that are sensitive to the cognitive states of the learner. The Affective AutoTutor augments these rules to be sensitive to dynamic assessments of learners' affective states, specifically boredom, confusion, and frustration. The affective states are sensed by automatically monitoring interaction patterns, gross body movements, and facial features (D'Mello & Graesser, 2012a). The Affective AutoTutor responds with empathetic, encouraging, and motivational dialog-moves along with emotional displays. For example, the tutor might respond to mild boredom with, "This stuff can be kind of dull sometimes, so I'm gonna try and help you get through it. Let's go". The affective responses are accompanied by appropriate emotional facial expressions and emotionally modulated speech (e.g., synthesized empathy or encouragement).

The effectiveness of Affective AutoTutor over the original non-affective AutoTutor was tested in a between-subjects experiment where 84 learners were randomly assigned to two 30-minute learning sessions with either tutor (D'Mello, Lehman, Sullins et al., 2010). The results indicated that the affective tutor helped learning for low-domain knowledge learners during the second 30-minute learning session. The affective tutor was less effective at promoting learning for high-domain knowledge learners during the first 30-minute session. Importantly, learning gains increased from Session 1 to Session 2 with the affective tutor whereas they plateaued with the non-affective tutor. Learners who interacted with the affective tutor also demonstrated improved performance on a subsequent transfer test. A follow-up analysis indicated that learners' perceptions of how closely the computer tutors resembled human tutors increased across learning sessions, was related to the quality of tutor feedback, and was a powerful predictor of learning (D'Mello & Graesser, 2012c). The positive change in perceptions was greater for the affective tutor.

As a second example, consider UNC-ITSPOKE (Forbes-Riley & Litman, 2011), a speech-enabled ITS for physics with the capability to automatically detect and respond to learners' certainty/uncertainty in addition to the correctness/incorrectness of their spoken responses. Uncertainty detection was performed by extracting and analyzing the acoustic-prosodic features in learners' spoken responses along with lexical and dialog-based features. UNC-ITSPOKE responded to uncertainty when the learner was correct but uncertain about the response. This was taken to signal an impasse because the learner is unsure about the state of their knowledge, despite being correct. The actual response strategy involved launching explanation-based sub-dialogs that provided added instruction to remediate the uncertainty. This could involve additional follow-up questions (for more difficult content) or simply the assertion of the correct information with elaborated explanations (for easier content).

Forbes-Riley and Litman (2011) compared learning outcomes of 72 learners who were randomly assigned to receive adaptive responses to uncertainty (adaptive condition), no responses to uncertainty (non-adaptive control condition), or random responses to uncertainty (random control condition). In this later condition, the added tutorial content from the sub-dialogs was given for a random set of turns in order to control for the additional tutoring. The results indicated that the adaptive condition achieved slightly (but not significantly) higher learning outcomes than the random and non-adaptive control conditions. The findings revealed that it was perhaps not the presence or absence of adaptive responses to uncertainty, but the number of adaptive responses that correlated with learning outcomes.

EMERGING THEMES

Research at the intersection of emotions, learning, LA, and EDM, has typically focused on one-on-one learning with intelligent tutoring systems (Forbes-Riley & Litman, 2011; Woolf et al., 2009), educational games (Conati & Maclaren, 2009; Sabourin, Mott, & Lester, 2011), or interfaces that support basic competencies like reading, writing, text-diagram integration, and problem solving (D'Mello & Graesser, 2014a; D'Mello, Lehman, & Person, 2010; D'Mello & Mills, 2014). Although these basic lines of research are quite active, recent work has focused on analyzing affect across more expansive interaction contexts that more closely capture the broader sociocultural context surrounding learning. I briefly describe four themes of research to illustrate a few of the exciting developments.

Affect-Based Predictors of Attrition and Dropout

Early indicators of risk and early intervention systems are some of the "killer apps" of LA and EDM (Jayaprakash, Moody, Lauría, Regan, & Baron, 2014). Most fielded systems focus on academic performance data, demographics, and availability of financial assistance. These factors are undoubtedly important, but there are likely alternate factors that come into play. With this in mind, Aguiar, Ambrose, Chawla, Goodrich, and Brockman (2014) compared the predictive power of traditional academic and demographic features with features indicative of behavioural engagement in predicting dropout from an Introduction to Engineering Course. Their key finding was that behaviourally engaging with e-portfolios, measured by number of logins, number of artifacts submitted, and number of page hits, was a better predictor of dropout than models constructed from academic performance and demographics alone. Although affect was not directly measured in this study, behaviourally engaging with e-portfolios can be considered a sign of interest, which is a powerful motivating emotion.

Sentiment Analysis of Discussion Forums

Language communicates feelings. Hence, sentiment analysis and opinion mining techniques (Pang & Lee, 2008) have considerable potential to study how students' thoughts (expressed in written language) about a learning experience predicts relevant behaviours (most importantly attrition). In line with this, Wen, Yang, and Rosé (2014) applied sentiment analysis techniques on student posts on three Massive Open Online Courses (MOOCs). They observed a negative correlation between the ratio of positive to negative terms and dropout across time. More recently, Yang, Wen, Howley, Kraut, and Rosé (2015) developed methods to automatically identify discussion posts that were indicative of student confusion. They showed that confusion reduced the likelihood of retention, but this could be mitigated with confusion resolution and other supportive interventions.

Classroom Learning Analytics

Recent advances in sensing and signal processing technologies have made it possible to automatically model aspects of students' classroom experience that could previously only be obtained from self-reports and cumbersome human observations. For example, second-generation Kinects can detect whether the eyes or mouth are open, if a person is looking away, and if the mouth has moved, for up to six people at a time (Microsoft, 2015). In one pioneering study, Raca, Kidzinski, and Dillenbourg (2015) tracked students in a classroom using multiple cameras affixed around the blackboard area. Computer vision techniques were used for head detection and head-pose estimation, which were then used to train a detector of student attention (validated via self-reports). This emerging area, related to the field of multimodal learning analytics (Blikstein, 2013), is poised for considerable progress in years to come.

Teacher Analytics

Teachers should not be left out of the loop since teacher practices are known to influence student affect and engagement. Unfortunately, quantifying teacher instructional practices relies on live observations in classrooms (e.g., Nystrand, 1997), which makes the research difficult to scale. To address this, researchers have begun to develop methods for automatic analysis of teacher instructional practices. In a pioneering study, Wang, Miller, and Cortina (2013) recorded classroom audio in 1st to 3rd grade math classes and developed automatic methods to predict the level of discussions in these classes. This work was recently expanded to analyze several additional instructional activities (lecturing, small group work, supervised seatwork, question/answer, and procedures and directions) in larger samples of middle-school literature and language-arts classes using teacher audio alone (Donnelly et al., 2016a) or a combination of teacher and classroom audio (Donnelly et al., 2016b). Blanchard et al. (2016) used teacher audio to automatically detect teacher questions, achieving a .85 correlation with the proportion of human-coded questions. The next step in this line of work is to use information on what teachers are doing to contextualize how students are feeling, which in turn influences what they think, do, and learn.

FUTURE THEMES

Let me end by briefly highlighting some potential future themes of research. One promising area of research involves a detailed analysis of the emotional experience of learners and communities of learners across the extended time scale of a traditional course, a flipped-course, or a MOOC (Dillon et al., 2016). A second involves the study of emotion regulation during learning, especially how LA/EDM methods can be used to identify different regulatory strategies (Gross, 2008), so that the more beneficial ones can be engendered (e.g., Strain & D'Mello, 2014). A third would jointly consider emotion alongside attentional states of mindfulness, mind wandering, and how emotion-attention blends like the "flow experience" (Csikszentmihalyi, 1990) emerge and manifest in the body and in behaviour. A fourth addresses how the so-called "non-cognitive" (Farrington et al., 2012) traits like grit, self-control, and diligence modulate learner emotions and efforts to regulate them (e.g., Galla et al., 2014). A fifth would monitor emotions of groups of learners during collaborative learning and collaborative problem solving (Ringeval, Sonderegger, Sauer, & Lalanne, 2013) given the importance of collaboration as a critical 21st century skill (OECD, 2015).

Finally, quoting William James's classic 1884 treatise on emotion: "The most important part of my environment is my fellow-man. The consciousness of his attitude towards me is the perception that normally unlocks most of my shames and indignations and fears" (p. 195). Research to date has mainly focused on the achievement, epistemic, and topic emotions. However, an analysis of learning in the sociocultural context in which it is situated must adequately address the social emotions of pride, shame, guilt, jealousy, envy, and so on. This is both a future theme and a grand research challenge.

CONCLUSION

Learning is not a cold intellectual activity; it is punctuated with emotion. The emotions are not merely decorative, they have agency. But emotion is a complex phenomenon with multiple components that dynamically unfold across multiple time scales. And despite great strides in the fields of affective sciences and affective neuroscience, we know little about emotions, and even less on emotions during learning. This is certainly not to imply that we should refrain from modelling emotion until there is more theoretical clarity. Quite the opposite. It simply means that we need to be mindful of what we are modelling when we say we are modelling emotion. We also need to embrace, rather than dilute, the complexity and ambiguity inherent in emotion. If anything, the discovery-oriented, data-driven, analytic methods of LA and EDM, along with an emphasis on real-world data collection, has the unique potential to advance both the science of learning and the science of emotion. It all begins by incorporating emotion into the analysis of learning.

ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation (NSF) (DRL 1108845 and IIS 1523091), the Bill & Melinda Gates Foundation, and the Institute for Educational Sciences (R305A130030). Any opinions, findings and conclusions, or recommendations expressed in this paper are those of the author and do not necessarily reflect the views of the funding agencies.

REFERENCES

Aguiar, E., Ambrose, G. A. A., Chawla, N. V., Goodrich, V., & Brockman, J. (2014). Engagement vs. performance: Using electronic portfolios to predict first semester engineering student persistence. *Journal of Learning Analytics*, 1(3), 7–33.

Ai, H., Litman, D. J., Forbes-Riley, K., Rotaru, M., Tetreault, J. R., & Pur, A. (2006). Using system and user performance features to improve emotion detection in spoken tutoring dialogs. *Proceedings of the 9th International Conference on Spoken Language Processing* (Interspeech 2006).

Arroyo, I., Woolf, B., Cooper, D., Burleson, W., Muldner, K., & Christopherson, R. (2009). Emotion sensors go to school. In V. Dimitrova, R. Mizoguchi, B. Du Boulay, & A. Graesser (Eds.), *Proceedings of the 14th International Conference on Artificial Intelligence in Education* (pp. 17–24). Amsterdam: IOS Press.

Baker, R., & Ocumpaugh, J. (2015). Interaction-based affect detection in educational software. In R. Calvo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing* (pp. 233–245). New York: Oxford University Press.

Barth, C. M., & Funke, J. (2010). Negative affective environments improve complex solving performance. *Cognition and Emotion*, 24(7), 1259–1268. doi:10.1080/02699930903223766

Blanchard, N., Donnelly, P., Olney, A. M., Samei, B., Ward, B., Sun, X., . . . D'Mello, S. K. (2016). Identifying teacher questions using automatic speech recognition in live classrooms. *Proceedings of the 17th Annual SIGdial Meeting on Discourse and Dialogue* (SIGDIAL 2016) (pp. 191–201). Association for Computational Linguistics.

Blikstein, P. (2013). Multimodal learning analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium. New York: ACM.

Bosch, N., Chen, H., Baker, R., Shute, V., & D'Mello, S. K. (2015). Accuracy vs. availability heuristic in multimodal affect detection in the wild. *Proceedings of the 17th ACM International Conference on Multimodal Interaction* (ICMI 2015). New York: ACM.

Bosch, N., & D'Mello, S. K. (in press). The affective experience of novice computer programmers. *International Journal of Artificial Intelligence in Education.*

Bosch, N., D'Mello, S., Baker, R., Ocumpaugh, J., & Shute, V. (2016). Using video to automatically detect learner affect in computer-enabled classrooms. *ACM Transactions on Interactive Intelligent Systems*, 6(2), doi:10.1145/2946837.

Calvo, R. A., & D'Mello, S. K. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1), 18–37. doi:10.1109/T-AFFC.2010.1

Clore, G. L., & Huntsinger, J. R. (2007). How emotions inform judgment and regulate thought. *Trends in Cognitive Sciences*, 11(9), 393–399. doi:10.1016/j.tics.2007.08.005

Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3), 267–303.

Corbett, A., & Anderson, J. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: Harper & Row.

D'Mello, S. K. (2016). On the influence of an iterative affect annotation approach on inter-observer and self-observer reliability. *IEEE Transactions on Affective Computing*, 7(2), 136–149.

D'Mello, S. K., Blanchard, N., Baker, R., Ocumpaugh, J., & Brawner, K. (2014). I feel your pain: A selective review of affect-sensitive instructional strategies. In R. Sottilare, A. Graesser, X. Hu, & B. Goldberg (Eds.), *Design recommendations for adaptive intelligent tutoring systems: Adaptive instructional strategies* (Vol. 2, pp. 35–48). Orlando, FL: US Army Research Laboratory.

D'Mello, S., Craig, S., Sullins, J., & Graesser, A. (2006). Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education*, 16(1), 3–28.

D'Mello, S., & Graesser, A. (2012a). AutoTutor and Affective AutoTutor: Learning by talking with cognitively and emotionally intelligent computers that talk back. *ACM Transactions on Interactive Intelligent Systems*, 2(4), 23:22–23:39.

D'Mello, S., & Graesser, A. (2012b). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145–157. doi:10.1016/j.learninstruc.2011.10.001

D'Mello, S., & Graesser, A. (2012c). Malleability of students' perceptions of an affect-sensitive tutor and its influence on learning. In G. Youngblood & P. McCarthy (Eds.), *Proceedings of the 25th Florida Artificial Intelligence Research Society Conference* (pp. 432–437). Menlo Park, CA: AAAI Press.

D'Mello, S., & Graesser, A. (2014a). Inducing and tracking confusion and cognitive disequilibrium with breakdown scenarios. *Acta Psychologica*, 151, 106–116.

D'Mello, S. K., & Graesser, A. C. (2014b). Confusion. In R. Pekrun & L. Linnenbrink-Garcia (Eds.), *International handbook of emotions in education* (pp. 289–310). New York: Routledge.

D'Mello, S. K., & Kory, J. (2015). A review and meta-analysis of multimodal affect detection systems. *ACM Computing Surveys*, 47(3), 43:41–43:46.

D'Mello, S., Lehman, B., & Person, N. (2010). Monitoring affect states during effortful problem solving activities. *International Journal of Artificial Intelligence in Education*, 20(4), 361–389.

D'Mello, S., Lehman, B., Sullins, J., Daigle, R., Combs, R., Vogt, K., . . . Graesser, A. (2010). A time for emoting: When affect-sensitivity is and isn't effective at promoting deep learning. In J. Kay & V. Aleven (Eds.), *Proceedings of the 10th International Conference on Intelligent Tutoring Systems* (pp. 245–254). Berlin/Heidelberg: Springer.

D'Mello, S., & Mills, C. (2014). Emotions while writing about emotional and non-emotional topics. *Motivation and Emotion*, 38(1), 140–156.

D'Mello, S., Taylor, R. S., & Graesser, A. (2007). Monitoring affective trajectories during complex learning. In D. McNamara & J. Trafton (Eds.), *Proceedings of the 29th Annual Cognitive Science Society* (pp. 203–208). Austin, TX: Cognitive Science Society.

Darwin, C. (1872). *The expression of the emotions in man and animals.* London: John Murray.

Dillon, J., Bosch, N., Chetlur, M., Wanigasekara, N., Ambrose, G. A., Sengupta, B., & D'Mello, S. K. (2016). Student emotion, co-occurrence, and dropout in a MOOC context. In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (EDM 2016) (pp. 353–357). International Educational Data Mining Society.

Donnelly, P., Blanchard, N., Samei, B., Olney, A. M., Sun, X., Ward, B., . . . D'Mello, S. K. (2016a). Automatic teacher modeling from live classroom audio. In L. Aroyo, S. D'Mello, J. Vassileva, & J. Blustein (Eds.), *Proceedings of the 2016 ACM International Conference on User Modeling, Adaptation, & Personalization* (ACM UMAP 2016) (pp. 45–53). New York: ACM.

Donnelly, P., Blanchard, N., Samei, B., Olney, A. M., Sun, X., Ward, B., . . . D'Mello, S. K. (2016b). Multi-sensor modeling of teacher instructional segments in live classrooms. *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (ICMI 2016). New York: ACM.

Ekman, P., & Friesen, W. (1978). *The facial action coding system: A technique for the measurement of facial movement.* Palo Alto, CA: Consulting Psychologists Press.

Emotient. (2014). FACET: Facial expression recognition software.

Farrington, C. A., Roderick, M., Allensworth, E., Nagaoka, J., Keyes, T. S., Johnson, D. W., & Beechum, N. O. (2012). *Teaching adolescents to become learners: The role of noncognitive factors in shaping school performance: A critical literature review*. Chicago, IL: University of Chicago Consortium on Chicago School Research.

Forbes-Riley, K., & Litman, D. J. (2011). Benefits and challenges of real-time uncertainty detection and adaptation in a spoken dialogue computer tutor. *Speech Communication*, 53(9–10), 1115–1136. doi:10.1016/j.specom.2011.02.006

Fredrickson, B., & Branigan, C. (2005). Positive emotions broaden the scope of attention and thought-action repertoires. *Cognition & Emotion*, 19(3), 313–332. doi:10.1080/02699930441000238

Galla, B. M., Plummer, B. D., White, R. E., Meketon, D., D'Mello, S. K., & Duckworth, A. L. (2014). The academic diligence task (ADT): Assessing individual differences in effort on tedious but important schoolwork. *Contemporary Educational Psychology*, 39(4), 314–325.

Graesser, A., Chipman, P., Haynes, B., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612–618. doi:10.1109/TE.2005.856149

Gross, J. (2008). Emotion regulation. In M. Lewis, J. Haviland-Jones, & L. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 497–512). New York: Guilford.

Hess, U. (2003). Now you see it, now you don't - the confusing case of confusion as an emotion: Commentary on Rozin and Cohen (2003). *Emotion*, 3(1), 76–80.

Isen, A. (2008). Some ways in which positive affect influences decision making and problem solving. In M. Lewis, J. Haviland-Jones, & L. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 548–573). New York: Guilford.

Izard, C. (2010). The many meanings/aspects of emotion: Definitions, functions, activation, and regulation. *Emotion Review*, 2(4), 363–370. doi:10.1177/1754073910374661

James, W. (1884). What is an emotion? *Mind*, 9, 188–205.

Jayaprakash, S. M., Moody, E. W., Lauría, E. J., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, 1(1), 6–47.

Kai, S., Paquette, L., Baker, R., Bosch, N., D'Mello, S., Ocumpaugh, J., . . . Ventura, M. (2015). Comparison of face-based and interaction-based affect detectors in physics playground. In C. Romero, M. Pechenizkiy, J. Boticario, & O. Santos (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM 2015) (pp. 77–84). International Educational Data Mining Society.

Kory, J., D'Mello, S. K., & Olney, A. (2015). Motion tracker: Camera-based monitoring of bodily movements using motion silhouettes. *PloS ONE*, 10(6), doi:10.1371/journal.pone.0130293.

Lench, H. C., Bench, S. W., & Flores, S. A. (2013). Searching for evidence, not a war: Reply to Lindquist, Siegel, Quigley, and Barrett (2013). *Psychological Bulletin*, 113(1), 264–268.

Lewis, M. D. (2005). Bridging emotion theory and neurobiology through dynamic systems modeling. *Behavioral and Brain Sciences*, 28(2), 169–245.

Lindquist, K. A., Siegel, E. H., Quigley, K. S., & Barrett, L. F. (2013). The hundred-year emotion war: Are emotions natural kinds or psychological constructions? Comment on Lench, Flores, and Bench (2011). *Psychological Bulletin*, 139(1), 264–268.

McKeown, G. J., & Sneddon, I. (2014). Modeling continuous self-report measures of perceived emotion using generalized additive mixed models. *Psychological Methods*, 19(1), 155–174.

Mesquita, B., & Boiger, M. (2014). Emotions in context: A sociodynamic model of emotions. *Emotion Review*, 6(4), 298–302.

Microsoft. (2015). Kinect for Windows SDK MSDN. Retrieved 5 August 2015 from https://msdn.microsoft.com/en-us/library/dn799271.aspx

Moors, A. (2014). Flavors of appraisal theories of emotion. *Emotion Review*, 6(4), 303–307.

Nystrand, M. (1997). Opening dialogue: Understanding the dynamics of language and learning in the English classroom. *Language and Literacy Series*. New York: Teachers College Press.

Ocumpaugh, J., Baker, R. S., & Rodrigo, M. M. T. (2012). Baker-Rodrigo Observation Method Protocol (BROMP) 1.0. Training Manual version 1.0. New York.

OECD (Organisation for Economic Co-operation and Development). (2015). PISA 2015 collaborative problem solving framework.

Ohman, A., & Soares, J. (1994). Unconscious anxiety: Phobic responses to masked stimuli. *Journal of Abnormal Psychology*, *103*(2), 231–240.

Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, *2*(1–2), 1–135.

Pardos, Z., Baker, R. S. J. d., San Pedro, M. O. C. Z., & Gowda, S. M. (2013). Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes. In D. Suthers, K. Verbert, E. Duval, & X. Ochoa (Eds.), *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 117–124). New York: ACM.

Parkinson, B., Fischer, A. H., & Manstead, A. S. (2004). *Emotion in social relations: Cultural, group, and interpersonal processes*. Psychology Press.

Pekrun, R., & Stephens, E. J. (2011). Academic emotions. In K. Harris, S. Graham, T. Urdan, S. Graham, J. Royer, & M. Zeidner (Eds.), *APA educational psychology handbook, Vol 2: Individual differences and cultural and contextual factors* (pp. 3–31). Washington, DC: American Psychological Association.

Porayska-Pomsta, K., Mavrikis, M., D'Mello, S. K., Conati, C., & Baker, R. (2013). Knowledge elicitation methods for affect modelling in education. *International Journal of Artificial Intelligence in Education*, 22, 107–140.

Raca, M., Kidzinski, L., & Dillenbourg, P. (2015). Translating head motion into attention: Towards processing of student's body language. *Proceedings of the 8th International Conference on Educational Data Mining*. International Educational Data Mining Society.

Razzaq, L., Feng, M., Nuzzo-Jones, G., Heffernan, N. T., Koedinger, K. R., Junker, B., . . . Choksey, S. (2005). The assistment project: Blending assessment and assisting. In C. Loi & G. McCalla (Eds.), *Proceedings of the 12th International Conference on Artificial Intelligence in Education* (pp. 555–562). Amsterdam: IOS Press.

Ringeval, F., Sonderegger, A., Sauer, J., & Lalanne, D. (2013). Introducing the RECOLA multimodal corpus of remote collaborative and affective interactions. *Proceedings of the 2nd International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space (EmoSPACE) in conjunction with the 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*. Washington, DC: IEEE.

Rosenberg, E. (1998). Levels of analysis and the organization of affect. *Review of General Psychology*, 2(3), 247–270. doi:10.1037//1089-2680.2.3.247

Rozin, P., & Cohen, A. (2003). High frequency of facial expressions corresponding to confusion, concentration, and worry in an analysis of maturally occurring facial expressions of Americans. *Emotion*, 3, 68–75.

Sabourin, J., Mott, B., & Lester, J. (2011). Modeling learner affect with theoretically grounded dynamic Bayesian networks. In S. D'Mello, A. Graesser, B. Schuller, & J. Martin (Eds.), *Proceedings of the 4th International Conference on Affective Computing and Intelligent Interaction* (pp. 286–295). Berlin/Heidelberg: Springer-Verlag.

San Pedro, M., Baker, R. S., Bowers, A. J., & Heffernan, N. T. (2013). Predicting college enrollment from student interaction with an intelligent tutoring system in middle school. In S. D'Mello, R. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM 2013) (pp. 177–184). International Educational Data Mining Society.

Schwarz, N. (2012). Feelings-as-information theory. In P. Van Lange, A. Kruglanski, & T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 289–308). Thousand Oaks, CA: Sage.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013). Assessment and learning of qualitative physics in Newton's playground. *The Journal of Educational Research*, 106(6), 423–430.

Sinha, T., Jermann, P., Li, N., & Dillenbourg, P. (2014). Your click decides your fate: Inferring information processing and attrition behavior from MOOC video clickstream interactions. Paper presented at the Empirical Methods in Natural Language Processing: Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses. http://www.aclweb.org/anthology/W14-4102

Strain, A., & D'Mello, S. (2014). Affect regulation during learning: The enhancing effect of cognitive reappraisal. *Applied Cognitive Psychology*, 29(1), 1–19. doi:10.1002/acp.3049

Tracy, J. L. (2014). An evolutionary approach to understanding distinct emotions. *Emotion Review*, 6(4), 308–312.

VanLehn, K., Siler, S., Murray, C., Yamauchi, T., & Baggett, W. (2003). Why do only some events cause learning during human tutoring? *Cognition and Instruction*, *21*(3), 209–249. doi:10.1207/S1532690XCI2103_01

Wang, Z., Miller, K., & Cortina, K. (2013). Using the LENA in teacher training: Promoting student involvement through automated feedback. *Unterrichtswissenschaft*, 4, 290–305.

Wen, M., Yang, D., & Rosé, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, S. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM 2014), 4–7 July, London, UK (pp. 130–137): International Educational Data Mining Society. http://www.cs.cmu.edu/~mwen/papers/edm2014-camera-ready.pdf

Woolf, B., Burleson, W., Arroyo, I., Dragon, T., Cooper, D., & Picard, R. (2009). Affect-aware tutors: Recognizing and responding to student affect. *International Journal of Learning Technology*, 4(3/4), 129–163.

Yang, D., Wen, M., Howley, I., Kraut, R., & Rosé, C. (2015). Exploring the effect of confusion in discussion forums of massive open online courses. *Proceedings of the 2nd ACM conference on Learning@Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 121–130). New York: ACM. doi:10.1145/2724660.2724677

Zeng, Z., Pantic, M., Roisman, G., & Huang, T. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 39–58.

Chapter 11: Multimodal Learning Analytics

Xavier Ochoa

Escuela Superior Politécnica del Litoral, Ecuador
DOI: 10.18608/hla17.011

ABSTRACT

This chapter presents a different way to approach learning analytics (LA) praxis through the capture, fusion, and analysis of complementary sources of learning traces to obtain a more robust and less uncertain understanding of the learning process. The sources or modalities in multimodal learning analytics (MLA) include the traditional log-file data captured by online systems, but also learning artifacts and more natural human signals such as gestures, gaze, speech, or writing. The current state-of-the-art of MLA is discussed and classified according to its modalities and the learning settings where it is usually applied. This chapter concludes with a discussion of emerging issues for practitioners in multimodal techniques.

Keywords: Audio, video, data fusion, multisensor

In its origins, the focus of the field of learning analytics (LA) was the study of the actions that students perform while using some sort of digital tool. These digital tools being learning management systems (LMSs; Arnold & Pistilli, 2012), intelligent tutoring systems (ITSs; Crossley, Roscoe, & McNamara, 2013), massive open online courses (MOOCs; Kizilcec, Piech, & Schneider, 2013), educational video games (Serrano-Laguna & Fernández-Manjón, 2014), or other types of systems that use a computer as an active component in the learning process. On the other hand, comparatively less LA research or practice has been conducted in other learning contexts, such as face-to-face lectures or study groups, where computers are not present or have only an auxiliary, not-defined role. This bias towards computer-based learning contexts is well explained by the basic requirement of any type of LA study or system: the existence of learning traces (Siemens, 2013).

Computer-based learning systems, even if not initially designed with analytics in mind, tend to capture automatically, in fine-grained detail, the interactions with their users. The data describing these interactions is stored in many forms; for example, log-files or word-processor documents that can be later mined to extract the traces to be analyzed. The relative abundance of readily available data and the low technical barriers to process it make computer-based learning systems the ideal place to conduct R&D for LA. On the contrary, in learning contexts where computers are not used, the actions of learners are not automatically captured. Even if some learning artifacts exist, such as student-produced physical documents, they need to be converted before they can be processed. Without traces to analyze, computational models and tools used traditionally in LA are not applicable.

The existence of this bias towards computer-based learning contexts could produce a streetlight effect (Freedman, 2010) in LA. This effect takes its name from a joke in which a man loses his house keys and searches for them under a streetlight even though he lost them in the park. A police officer watching the scene asks why he is searching on the street then, to which the man responds, "because the light is better over here." The streetlight effect means looking for solutions where it is easy to search, not where the real solutions might be. The case can be made for early LA research trying to understand and optimize the learning process by looking only at computer-based contexts but ignoring real-world environments where a large part of the process still happens. Even learners' actions that cannot be logged in computer-based systems are usually ignored. For example, the information about a student looking confused when presented with a problem in an ITS or yawning while watching an online lecture is not considered in traditional LA research. To diminish the streetlight effect, researchers are now focusing on how to collect fine-grained learning traces from real-world learning contexts automatically, making the analysis of a face-to-face lecture as feasible as the analysis of a MOOC session. More contemporary works on LA explore the new sources of data apart from traditional log-files: student-generated texts (Simsek et al., 2015), eye-tracking information (Vatrapu, Reimann, Bull, & Johnson, 2013) and classroom configuration (Almeda, Scupelli, Baker, Weber, & Fisher, 2014) to name a few. The combination of these different sources of learning traces into a single analysis is the main objective of multimodal learning analytics (MLA).

MLA is a subfield that attempts to incorporate different sources of learning traces into LA research and practice by focusing on understanding and optimizing learning in digital and real-world scenarios where the interactions are not necessarily mediated through a computer or digital device (Blikstein, 2013). In MLA, learning traces are combined from not only extracted from log-files or digital documents but from recorded video and audio, pen strokes, position tracking devices, biosensors, and any other modality that could be useful to understand or measure the learning process. Moreover, in MLA, the traces extracted from different modalities are combined to provide a more comprehensive view of the actions and the internal state of the learner.

The idea of using different modalities to study learning, while new in the context of LA, is common in traditional experimental educational research. Adding a human observer, which is by nature a multimodal sensor, into a real-world learning context is the usual way in which learning in-the-wild has been studied (Gall, Borg, & Gall, 1996). Technologies such as video and audio recording and tagging tools have made this observation less intrusive and more quantifiable (Cobb et al., 2003; Lund, 2007). The main problem with the traditional educational research approach is that the data collection and analysis, due to their manual nature, are very costly and do not scale. The data collection needs to be limited in both size and time and data analysis results are not available fast enough to be useful for the learners being studied. If different modalities could be recorded and learning traces could be automatically extracted from them, LA tools could be used to provide a continuous real-time feedback loop to improve learning as it is happening.

As would be expected, extracting learning traces from raw multimodal recordings is not trivial. Techniques developed in computer vision, speech processing, sketch recognition and other computer science fields must be guided by the current learning theories provided by learning science, educational research, and behavioural science. Given its complexity, the MLA subfield is relatively young and unexplored. However, initial studies and early interdisciplinary co-operation between researchers have produced positive results (Scherer, Worsley, & Morency, 2012; Morency, Oviatt, Scherer, Weibel, & Worsley, 2013; Ochoa, Worsley, Chiluiza, & Luz, 2014, Markaki, Lund, & Sanchez, 2015). This chapter is an initial guide for researchers and practitioners who want to explore this subfield. First, the main modalities used in MLA research will be presented, analyzed, and exemplified. Second, the real-world settings where MLA has been applied are studied and classified according to their main modalities. Finally, several unresolved issues important for MLA research and practice are discussed.

MODALITIES AND MEDIA

In its communication theory definition, multimodality refers to the use of diverse modes of communication (textual, aural, linguistic, spatial, visual, et cetera) to interchange information and meaning between individuals (Kress & Van Leeuwen, 2001). The media – movies, books, web pages, or even air – are the physical or digital substrate where a communication mode can be encoded. Each mode can be expressed through one or several media. For example, speech can be encoded as variations of pressure in the air (in a face-to-face dialog), as variations of magnetic orientation on a tape (in a cassette recording), or as variations of digital numbers (in an MP3 file). As well, the same medium can be used to transmit several modes. For example, a video recording can contain information about body language (posture), emotions (face expression), and tools used (actions).

By its own nature, learning is often multimodal (Jewitt, 2006). A human being can learn by reading a book, listening to a professor, watching a procedure, using physical or digital tools, and any other mode of human communication where relatively complex information can be encoded. The learning process also uses several feedback loops – for example, a student nodding when the instructor asks if the lesson was understood, or the emphasis of the instructor's voice while explaining a topic. These feedback modes usually encode simpler information but are critical for the process. If learning is to be analyzed, understood, and optimized, traces of the interactions occurring in each of the relevant modes should be obtained. MLA focuses on extracting these traces from the different modes of communication while being agnostic of the medium where those modes are encoded or recorded.

The following subsections present the state-of-the-art on the capture and trace-extraction for the most common modalities used in MLA research. For each modality, a brief definition is presented, together with a discussion of its importance to understanding the learning process, a list of most common methods of capture and recording, and examples of where they have been used. This is not a comprehensive list of all the modes relevant for learning, only those used successfully in MLA studies.

Gaze

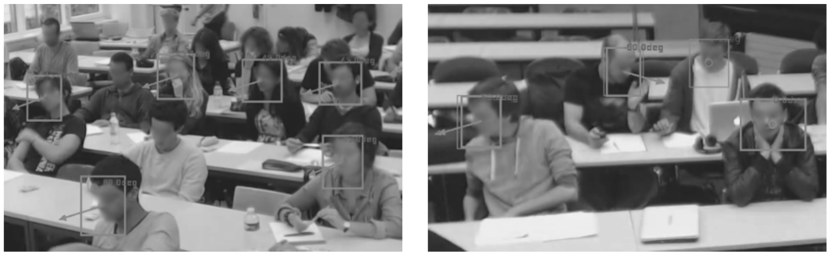
Humans tend to look directly at what draws their attention. As such, the direction of the gaze of an individual is a proxy indicator of the direction of his or her attention (Frischen, Bayliss, & Tipper, 2007). Attention is an indispensable requirement for learning (Kruschke, 2003). Paying attention to a signal helps the individual to capture its information and store the relevant parts in long-term memory. While gaze is not the only proxy to estimate attention and is not error-free, it is commonly used in educational practice. For example, a trained instructor can assess the level of attention of a whole classroom by surveying the gaze of the students; an observer can determine a participant's level of attention in a discussion by tracking the re-direction of the gaze from speaker to speaker.

The importance of gaze has been long identified by marketers, behavioural, and human–computer interaction researchers. Eye-tracking studies are common to determine the effectiveness of advertising (Krugman, Fox, Fletcher, Fischer, & Rojas, 1994), help with the early diagnosis of autism (Boraston & Blakemore, 2007), and the effectiveness of computer interfaces (Poole & Ball, 2006). However, the main methods for recording gaze in these studies, using monitor fixed eye-trackers or special eye-tracking glasses, are too intrusive and costly to be widely deployed in learning settings. The current medium of choice for gaze capturing in MLA is video recordings (Raca & Dillenbourg, 2013). A camera, or an array of cameras, is positioned to record the head and eyes of the subject(s). Then, computer vision techniques, such as those presented in Lin, Lin, Lin, and Lee (2013), are used to extract the gaze direction information from the video recording. The main aspects that need to be controlled to obtain the relative gaze direction in the recording are face resolution and avoiding occlusion from objects or other individuals in the setting (Raca & Dillenbourg, 2013). Information about the position of the cameras in the learning setting must also be recorded to calculate the absolute gaze direction.

MLA has several examples of gaze trace extraction. Raca and Dillenbourg (2013) estimate gaze direction from head orientation in video recordings of students sitting in a lecture using a part-based model (Figure 11.1). In this figure, student faces are automatically recognized (rectangle) and their gaze (arrow) is estimated based on a human face model. This information is then used to determine the focus of attention of individual students and compare it with self-reported attention. Raca and Dillenbourg found that the percentage of time students have the instructor in their field of vision is an important predictor of the level of attention reported. In a different learning setting, Echeverría, Avendaño, Chiluiza, Vásquez, and Ochoa (2014), also estimated gaze direction measuring head orientation by calculating the distance between eye centre points to nose tip point. This information was used to determine if students maintained eye contact with the audience during academic presentations.

Posture, Gestures, and Motion (Body Language)

Posture, gestures, and motion are three interrelated modes, jointly referred as body language, although each one could carry different types of information (Bull, 2013). Posture refers to the position that the body or part of the body adopts at a given moment in time. The posture of a learner could provide information about their internal state. For example, if a student is seated with the head resting on the desk, the instructor could infer that the student is tired or not interested in the lecture. In special cases, the posture adopted is related to the acquisition of skills. For example, students training in oral presentations are expected to use certain postures (hands and arms slightly open) rather than others (hands in the pockets). Gestures being learned do not indicate an internal state. Gestures are coordinated movements from different parts of the body, especially the head, arms, and hands to communicate a specific meaning. This non-verbal form of communication is usually conscious. It is used as a way to provide short feedback loops and alternative emphasizing channels in the learning process. For example, the instructor pointing to a specific part of



a) b)

Figure 11.1. Gaze estimation in a classroom setting (Raca, Tormey, & Dillenbourg, 2014).

the blackboard or a student raising his shoulders when confronted with a difficult question. Finally, motion is any change in body position not necessary to acquire a new posture or to perform a given gesture. This motion is often the result of unconscious body movements that reveal the inner state of the subject during the learning process; for example, erratic movements that signal nervousness or doubt.

Posture, gestures, and motion have been the modes most often studied in MLA due the relative ease in capturing video in real-world environments, together with the availability of low-cost 2-D and 3-D sensors and high-performing computer vision algorithms for feature extraction. While body language can be captured with high precision using accelerometers attached to different body parts (Mitra & Acharya, 2007) or using specialized tools (for example, a Wii Remote; Schlömer, Poppinga, Henze, & Boll, 2008), in practice using them is too invasive or foreign in most learning activities. The most common solution to capture motion is recording video of the subject and estimating posture, gestures, and motion. Any type of camera can be used as long as it can capture the relevant motion with enough resolution. The resolution needed depends on the type of feature extraction conducted with the video. For automatic extraction of human motion, the most common device used is the Microsoft Kinect (Zhang, 2012). Through a mixture of video and depth capture, Kinect is able to provide researchers with a reconstructed skeleton of the subject for each captured frame, which is ideal for capturing body postures and gestures. Newer versions of the Kinect sensor are also able to extract hand gestures (Vasquez, Vargas, & Sucar, 2015).

The most salient examples of the capture and processing of body language in MLA are the estimation of attention through upper-body relative movement delay in a classroom setting (Raca, Tormey, & Dillenbourg, 2014) and the posture and gesture analysis of a novice academic presenter towards the creation of an automated presentation tutor (Echeverría et al., 2014). Figure 11.2 presents the 23 different postures

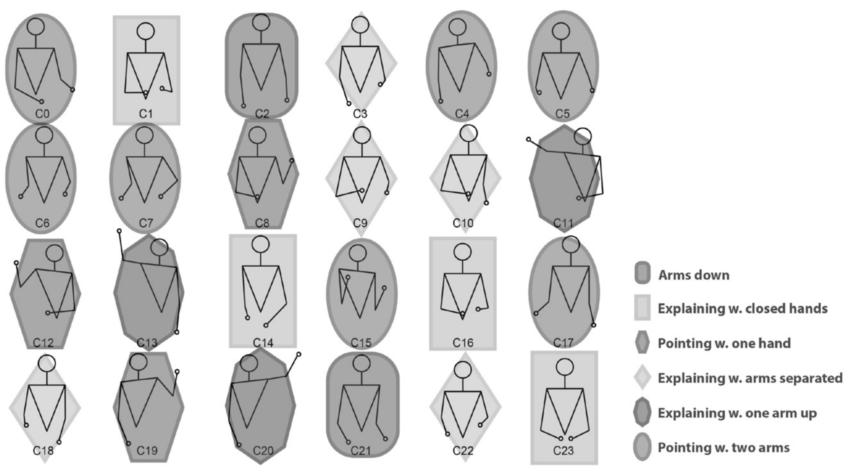


Figure 11.2. Clustered upper-body postures of real student presenters (Echeverría, Avendaño, Chiluiza, Vásquez, & Ochoa, 2014).

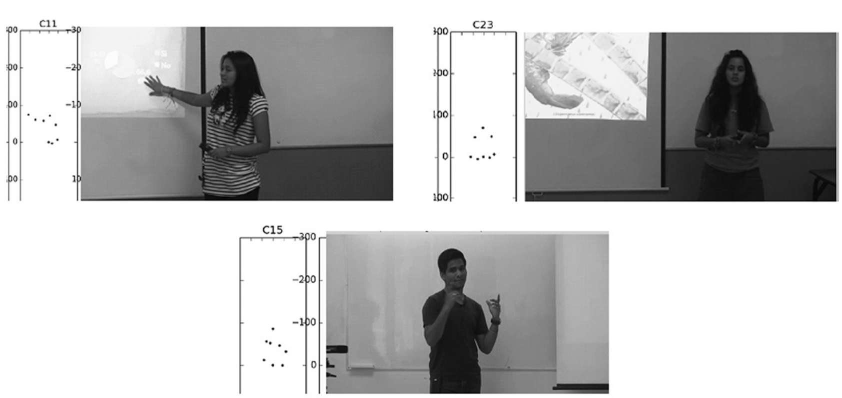


Figure 11.3. Actual postures classified according to prototype postures (Echeverría et al., 2014).

obtained from the analysis of Kinect data of students presenting their work. These 23 postures were classified into six body gestures (different colours) that could be considered good or bad for a presentation. Figure 11.3 presents real examples of these body gestures during actual presentations. The classification of the pose (above the Kinect points on the left) corresponds with what a human observer could interpret from the photo (on the right).

Other interesting examples in using gestures are Boncoddo et al. (2013), Alibali, Nathan, Fujimori, Stein, and Raudenbush, (2011), and Mazur-Palandre, Colletta, and Lund (2014). In the first, Boncoddo et al. (2013) captured the number of relevant gestures performed during the explanation of mathematical proofs and established the relation with the way students arrive at their answers. In the second, Alibali et al. (2011) classified the different gestures made by teachers during math classes and found relations between them. Finally, Mazur-Palandre et al. (2014) presented a study on the use of gestures by children when explaining procedures and instructions.

Actions

The action mode is very similar to the gesture and motion modes. Both are body movements usually captured by video recordings in MLA. However, actions are purposeful movements, usually involving the manipulation of a tool, that are usually learned. The type, sequence, or correctness of these actions can be used as indicators of the level of mastery that the learner has achieved in a given skill. For example, the order and security in which diverse tools are manipulated by a student in a lab can be used as a proxy to determine the understanding that the student has about a given procedure.

The main uses of action recording and analysis in MLA are in expertise estimation. In an engineering building activity, for example, the analysis of hand and wrist movement can determine the level of expertise (Worsley & Blikstein, 2014b). In mathematical problem solving, the percentage of time that a learner uses a calculator can be measured (Ochoa et al., 2013). Ochoa et al. (2013) tracked the position and angle of the calculator in problem-solving sessions (Figure 11.4). This position and angle (line) were then used to estimate which student was using the calculator during that specific frame in the video (intersection with the border of the image).

Facial Expressions

Also highly related to body language modes is the information gathered through facial expressions. The human face can communicate very complex mental states through relatively simple expressions. There has been a large body of successful research in the area of computer vision, trying to identify emotions automatically from facial expressions recorded in video (Mishra et al., 2015).

The main examples of using facial expressions in the field of LA are the works of Craig, D'Mello, Witherspoon, and Graesser (2008), and Worsley and Blikstein (2015b). Craig et al. (2008) automatically estimated the emotional states of students while using the AutoTutor system (Graesser, Chipman, Haynes, & Olney, 2005). Worsley and Blikstein (2015b) used similar techniques to discover emotional changes when students are confronted with different building exercises. Both studies discovered that a confused expression is a good indicator of the success of the learning process.

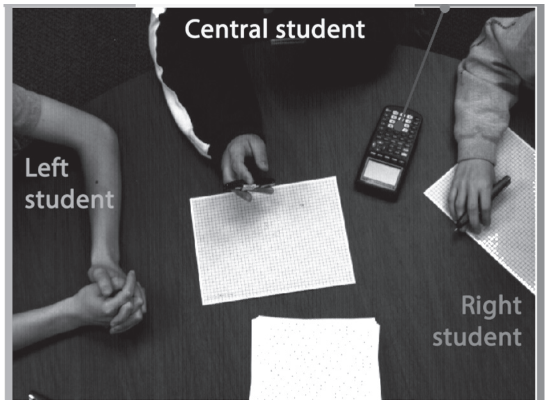


Figure 11.4. Determination of calculator use for expertise estimation (Ochoa et al., 2013).

Speech

The most common use of audio recordings in MLA is to capture traces of what the student is talking about or listening to. As the main and most complex form of communication among humans, speech is especially important in understanding the learning process. In the current practice of MLA, two main signals are extracted from audio recordings: what is being said and how it is being said. In the first approach, usually referred as speech recognition, the actual content of speech is extracted. The result of this analysis is a transcript that can be further processed using natural language processing (NLP) tools to establish what the subject is talking about. In the second approach, prosodic characteristics of the speech, such as intonation, tone, stress, and rhythm, are extracted. These characteristics can shed light on the internal state (security, emotional state, et cetera) or intention of the speaker (joke, sarcasm, et cetera). Speech recognition is heavily dependent on the language used, while prosodic characteristics are less sensible to language variations.

Audio is captured through microphones. While it seems easier to capture than video, the recording of audio of high enough quality to be processed is actually much more complicated. The type and spatial configuration of the microphones depend on the learning environment and what type of analysis will be conducted with the recorded signal. For example, if automatic speech recognition will be attempted, the microphone should be directional and be close to the subject's mouth. On the other hand, if only the detection of when somebody is talking is needed, an environmental microphone located in the middle of the room could be enough. The presence of noise and multiple signals not only prevents automatic feature extraction but will also degrade manual annotation. The most common technique used to improve recordings, when individual close-recording is not possible, is the use of microphone arrays that can not only reduce the noise but also determine the spatial origin of the audio.

Due to its importance, audio is also present in most MLA works to date. Different types of speech analysis have been used to establish the level of affinity of collaborative learning dialogues (Lubold & Pon-Harry, 2014), to evaluate the quality of oral presentations (Luzardo, Guamán, Chiluiza, Castells, & Ochoa, 2014), and to determine expertise in mathematics problem solving (Thompson, 2013).

Writing and Sketching

Two closely related modes are writing and sketching. They both use an instrument, most commonly a pen, to communicate complex thoughts. Using a pen is perhaps one of the first skills that students learn so using it to write and sketch is still a predominant activity in learning, especially at early stages. The most common information extracted from this mode is the transcript of what the student is saying, in the case of writing, or a structured representation of the sketches where information about their content can be inferred. However, capturing the process of writing and sketching through technological means opens the door to using information that human observers cannot easily detect, such as writing speed, rhythm, and pressure level. While their value for understanding learning is still not clear, there are indications that they could be good expertise predictors (Ochoa et al., 2013).

The recording instrument most commonly used to capture writing and sketching is a digital pen (Oviatt & Cohen, 2015). These pens are able to digitize the position, duration, and pressure of the strokes done on different surfaces. Once in digital form, this information can be used in LA tools. Alternatively, the widespread use of tablets in education (Clarke & Svanaes, 2014) also offers an opportunity to capture these modes easily, especially sketching.

In the realm of MLA, two works based on the math data corpus (Oviatt, Cohen, & Weibel, 2013) explored the contribution that writing and sketching modes could have in the prediction of expertise. Ochoa et al. (2013) extracted writing characteristics (stroke speed and length) and performed sketch recognition to determine the number of simple geometrical figures used. The results determined that speed of writing is highly correlated with level of expertise. Zhou, Hang, Oviatt, Yu, & Chen (2014) used classification systems based on writing and sketching characteristics to identify the expert in the group with 80% accuracy.

CONTEXTS

The main goal of MLA research is to extend the application of LA tools and methodologies to learning contexts that do not readily provide digital traces. One characteristic of these contexts is that the capture of more than one mode is necessary to understand the learning process. Table 11.1 presents a summary of the context studied in the current MLA literature with a detail of the modes used, the main learning aspects being explored in those contexts, and the works where those studies are conducted.

Table 11.1. Learning Contexts Studied by MLA

Contexts	Modes	Learning Aspects	Works
Lectures	Movement, Gaze, Gestures, Facial Expression, Speech	Attention, Question-Answer Interactions	Raca & Dillenbourg, 2013; Raca et al., 2014; Dominguez et al., 2015; D'Mello et al., 2015; Alibali et al., 2011
Oral Presentations	Posture, Movement, Gestures, Gaze, Speech, Digital Document	Skill Development, Feedback, Mental State	Luzardo et al., 2014; Echeverría et al., 2014; Chen et al., 2014; Leong et al., 2015; Schneider et al., 2015; Boncoddo et al., 2013
Problem-Solving	Movement, Actions, Speech, Writing, Sketching	Expertise Estimation	Ochoa et al., 2013; Luz, 2013; Thompson, 2013; Zhou et al., 2014
Construction Exercises	Gestures, Actions, Speech, Facial Expressions, Galvanic Skin Response	Novice vs. Expert Patterns	Worsley & Blikstein, 2013, 2014b, 2015a
Use of Intelligent Tutoring Systems	Digital Log Files, Facial Expressions, Speech	Relation between Emotions and Learning	Craig et al., 2008; D'Mello et al., 2008

Figure 11.5. Design of a multimodal recording device (MRD) to be used in lecture settings. MRD in the classroom (left) and MRD from student's point of view (right)

Lectures

Traditionally, lectures are the most common context associated with learning. While several aspects of this setting deserve study, MLA researchers to date have focused on automatically assessing the attention of students during the lecture. The seminal works of Raca and Dillenbourg (2013) and Raca et al. (2014) have explored the recording of video in the classroom and the automatic extraction of student movement and gaze from that recording. The results of these studies suggest that both modes are related to student attention, but there are other significant contributors, such as sitting position. Dominguez, Echeverría, Chiluiza, and Ochoa (2015) presented a novel, distributed way to capture video-, audio- and pen-based modes using a multimodal recording device (MRD). Figure 11.5 presents the design of such a device. The proximity of the device to the students reduces the risk of occlusion and increases the video and audio capture quality. Finally, D'Mello et al. (2015) produced diverse audio recordings in a lecture setting in order to evaluate question–answer interactions between instructor and students.

Oral Presentations

The skill of presenting an academic topic in front of an audience is frequently regarded as one of the soft-skills that higher-education students should acquire (Debnath et al., 2012). Several independent groups around the globe have recently started to build MLA systems able to help novice students correct bad-practices and gain mastery in oral presentations. Echeverría et al. (2014) and Luzardo et al. (2014) present different aspects of the same system that uses gesture, posture, movement, gaze, speech, and an analysis of the digital presentation files and is able to predict the grade that a human evaluator will give the student. Chen, Leong, Feng, and Lee (2014), analyzing the same data, were able to combine the different modalities in composite variables also used to predict the score. Schneider, Börner, van Rosmalen, and Specht (2015) also created a virtual presentation skill trainer utilizing Kinect to recognize postures and provide feedback in real-time.

Problem-Solving

Learning, especially in STEM subjects, frequently occurs at individual and group problem-solving sessions (Silver, 2013). The existence of the math data corpus (Oviatt et al. 2013), a set of multimodal recordings of groups of three high-school students solving math and geometry problems, catalyzed MLA research in this setting. The media provided in the dataset include frontal video recordings of each student, video recordings of the working table, audio recordings of each student, and general audio of the room. Additionally, students were equipped with digital pens. Ground truth is provided about the level of expertise of the students. Luz (2013), Thompson (2013), Ochoa et al. (2013), and Zhou et al. (2014) have all analyzed this dataset using diverse modes, concluding that all the modalities contributed to the determination of the level of expertise with a high level of accuracy (>70%).

Construction Exercises

The knowledge and skills required for engineering design and construction can be tested through small construction challenges (Householder & Hailey, 2012). The seminal works of Worsley and Blikstein (2013, 2014b, 2015a) explore, through multimodal analysis, the patterns of actions performed by experts and novices in the design and manual assembly of structures. The main modes used for the analysis were gestures, actions, speech, facial expression, and galvanic skin response. The combination of traces extracted from these modes reveals differences in the construction process that are helpful to identify the level of mastery in engineering design.

Use of Intelligent Tutoring Systems

ITSs are usually studied by traditional LA using log-files. However, video and audio of the learner have been captured to add new modes that complement the interaction data. The main modes extracted from the video are facial expression (Craig et al., 2008) and speech (D'Mello et al., 2008), which act as proxies for the learner's internal emotional state. Both are able to successfully detect emotional states such as boredom, confusion, and frustration in using the ITS.

SPECIFIC ISSUES

Once extracted, using multimodal traces in LA models and applications is similar to using different traces extracted from the same mode. However, MLA research and practice raise several specific issues when certain modalities are captured, processed, and analyzed. These issues remain open research areas, parallel to the technical extraction of traces from several modalities, but as important for the effective deployment of MLA solutions in the real-world.

Recording

Capturing interaction information in a digital tool is as easy and inexpensive as adding log statements in relevant parts of the code. These statements perform automatically, without requiring any involvement from the learner, in a transparent and generally reliable and error-free way. On the other hand, capturing media in the real-world requires the acquisition, installation, and use of recorders (cameras, microphones, digital pens, et cetera), turning the system on and off and monitoring it, and avoiding the degradation of the recording through occlusions, interference, or noise. Developing recording systems that work as effortlessly and efficiently as digital logging is one of the main barriers to the development of MLA. While this is an engineering problem, researchers should be aware of the feasibility and scalability of their solutions. One of the main proposals is to decentralize the recordings using inexpensive sensors that are always left on. If one or more recordings present problems, the general information could be reconstructed from the remaining working sensors.

Privacy

Capturing interaction information with digital tools already raises privacy concerns among students and instructors (Pardo & Siemens, 2014). The installation and use of recording systems that mimic "1984" levels of surveillance is bound to meet strong resistance. Informed consent forms could work for early research stages, but adopting MLA systems in the real-world would require a different, more creative approach. One of the most promising solutions in this area is transferring data ownership to the learner. Even if highly personal information is captured, privacy concerns are defused if the decision of what and when to share it remain in the control of the learner. This approach is similar to several quantified-self applications (Swan, 2013).

Integration

One question concerning the availability of large amounts of raw learning traces is how to combine them in order to produce useful information to understand and optimize the learning process. Traces extracted from different modes using different processes are bound to have very different characteristics. For example, the time granularity of the traces extracted from different modes can vary widely. Traces extracted from prosodic aspects of speech could change in tenths of a second while postures change more slowly. The level of certainty of the extracted traces can also be different. Speech recognition with high-quality recordings could reach 90% accuracy while emotional state detection from webcam sources could be in the low 70s. These difference do not prevent successful analysis, however, thoughtful design is required in order to prevent spurious results. Pioneering this line of research in MLA, Worsley and Blikstein (2014a) propose several fusion strategies based on the "bands of cognition" framework proposed by Newell (1994) and Anderson (2002) as an explanation for human cognition. The development of integration frameworks will benefit not only MLA but the whole LA community.

Impact on Learning

While the end-user tools and interventions based on multimodal learning analytics are similar to those based on monomodal analysis, the required usefulness of multimodal ones should be higher to justify the additional complexity of data acquisition. For example, a dashboard application based on data automatically captured by the LMS will be easier to accept than a similar dashboard that requires all classrooms be equipped with video cameras. The increased complexity should be accompanied by a larger positive impact on the learning process. The requirement of using multiple real-world signals to analyze learning should also come with the promise to provide more useful insights on the process and more measurable impacts on learners.

CONCLUSION

LA has revolutionized the approaches used to understand and optimize the learning process. However, its current bias towards studies and tools involving only computer-based learning contexts jeopardizes its applicability to learning in general. MLA is a subfield that seeks to integrate non-computer mediated learning contexts into the mainstream research and practice of LA.

This chapter presented the current-state-of-art in MLA. Modes as diverse as posture, speech, and sketching, alongside the more traditional modes of clickstream information and textual content, has been used to answer research questions and to build feedback systems in learning contexts. A mixture of computer science techniques and insights provided by educational and behavioural scientists enable the automatic evaluation of very diverse learning contexts, such as classrooms, study groups, and oral presentations.

As can be inferred from the list of research presented in this chapter, MLA is still a nascent field with a small but very active and open community of researchers. The existence of regular challenges and workshops, where multimodal datasets are freely shared and jointly analyzed with new designs ideas openly discussed, creates a research environment where new knowledge is generated rapidly.

While several issues still prevent MLA from becoming a mainstream practice, active research projects are exploring solutions to those issues, making the capture of multimodal learning traces cheaper, less invasive, and more automatic. Novel solutions born from the MLA community to handle privacy concerns, such as providing distributed recording and resting the ownership of the data with the learner, could one day be the norm for general LA practices.

Finally, the author wishes to invite LA researchers and practitioners to explore the use of multiple modalities in their own studies and tools. The MLA community will openly share its knowledge, data, code, and frameworks. Only the embrace of these different modalities will allow LA to have an impact in all the contexts where learning takes place.

REFERENCES

Alibali, M. W., Nathan, M. J., Fujimori, Y., Stein, N., & Raudenbush, S. (2011). Gestures in the mathematics classroom: What's the point? In N. Stein & S. Raudenbush (Eds.), *Developmental cognitive science goes to school* (pp. 219–234). New York: Routledge, Taylor & Francis.

Almeda, M. V., Scupelli, P., Baker, R. S., Weber, M., & Fisher, A. (2014). Clustering of design decisions in classroom visual displays. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 44–48). New York: ACM.

Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science*, *26*(1), 85–112.

Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 267–270). New York: ACM.

Blikstein, P. (2013). Multimodal learning analytics. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 102–106). New York: ACM.

Boraston, Z., & Blakemore, S.-J. (2007). The application of eye-tracking technology in the study of autism. *The Journal of Physiology*, 581(3), 893–898.

Bull, P. E. (2013). *Posture & Gesture*. Elsevier.

Boncoddo, R., Williams, C., Pier, E., Walkington, C., Alibali, M., Nathan, M., Dogan, M. & Waala, J. (2013). Gesture as a window to justification and proof. In M. V. Martinez & A. C. Superfine (Eds.), *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (PME-NA 35) 14–17 November 2013, Chicago, IL, USA (pp. 229–236). http://www.pmena.org/proceedings/

Chen, L., Leong, C. W., Feng, G., & Lee, C. M. (2014). Using multimodal cues to analyze MLA '14 oral presentation quality corpus: Presentation delivery and slides quality. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (MLA '14), 12–16 November 2014, Istanbul, Turkey (pp. 45–52). New York: ACM.

Clarke, B., & Svanaes, S. (2014, April 9). An updated literature review on the use of tablets in education. Family Kids and Youth. https://smartfuse.s3.amazonaws.com/mysandstorm.org/uploads/2014/05/T4S-Use-of-Tablets-in-Education.pdf

Cobb, P., Confrey, J., Lehrer, R., Schauble, L., & others. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.

Craig, S. D., D'Mello, S., Witherspoon, A., & Graesser, A. (2008). Emote aloud during learning with AutoTutor: Applying the facial action coding system to cognitive–affective states during learning. *Cognition and Emotion*, 22(5), 777–788.

Crossley, S. A., Roscoe, R. D., & McNamara, D. S. (2013). Using automatic scoring models to detect changes in student writing in an intelligent tutoring system. *Proceedings of the 26th Annual Florida Artificial Intelligence Research Society Conference* (FLAIRS-13), 20–22 May 2013, St. Pete Beach, FL, USA (pp. 208–213). Menlo Park, CA: The AAAI Press.

Debnath, M., Pandey, M., Chaplot, N., Gottimukkula, M. R., Tiwari, P. K., & Gupta, S. N. (2012). Role of soft skills in engineering education: Students' perceptions and feedback. In C. S. Nair, A. Patil, & P. Mertova (Eds.), *Enhancing learning and teaching through student feedback in engineering* (pp. 61–82). ScienceDirect. http://www.sciencedirect.com/science/book/9781843346456

D'Mello, S. K., Jackson, G. T., Craig, S. D., Morgan, B., Chipman, P., White, H., Person, N., Kort, B., el Kaliouby, R., Picard, R., & Graesser, A. C. (2008). AutoTutor detects and responds to learners' affective and cognitive states. Workshop on Emotional and Cognitive Issues in ITS, held in conjunction with the 9th International Conference on Intelligent Tutoring Systems (ITS 2008), 23–27 June 2008, Montreal, PQ, Canada. https://www.researchgate.net/publication/228673992_AutoTutor_detects_and_responds_to_learners_affective_and_cognitive_states

D'Mello, S., Olney, A., Blanchard, N., Samei, B., Sun, X., Ward, B., & Kelly, S. (2015). Multimodal capture of teacher–student interactions for automated dialogic analysis in live classrooms. *Proceedings of the 17th ACM International Conference on Multimodal Interaction* (ICMI '15), 9–13 November 2015, Seattle, WA, USA (pp. 557–566). New York: ACM.

Dominguez, F., Echeverría, V., Chiluiza, K., & Ochoa, X. (2015). Multimodal selfies: Designing a multimodal recording device for students in traditional classrooms. *Proceedings of the 17th ACM International Conference on Multimodal Interaction* (ICMI '15), 9–13 November 2015, Seattle, WA, USA (pp. 567–574). New York: ACM.

Echeverría, V., Avendaño, A., Chiluiza, K., Vásquez, A., & Ochoa, X. (2014). Presentation skills estimation based on video and Kinect data analysis. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (MLA '14), 12–16 November 2014, Istanbul, Turkey (pp. 53–60). New York: ACM.

Freedman, D. H. (2010, December 10). Why scientific studies are so often wrong: The streetlight effect. *Discover Magazine*, *26*. http://discovermagazine.com/2010/jul-aug/29-why-scientific-studies-often-wrong-streetlight-effect

Frischen, A., Bayliss, A. P., & Tipper, S. P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724.

Gall, M. D., Borg, W. R., & Gall, J. P. (1996). *Educational research: An introduction*. Longman Publishing.

Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005). AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, *48*(4), 612–618.

Householder, D. L., & Hailey, C. E. (2012). *Incorporating engineering design challenges into STEM courses*. National Center for Engineering and Technology Education. http://ncete.org/flash/pdfs/NCETECaucusReport.pdf

Jewitt, C. (2006). *Technology, literacy and learning: A multimodal approach*. Psychology Press.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 170–179). New York: ACM.

Kress, G., & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Edward Arnold.

Krugman, D. M., Fox, R. J., Fletcher, J. E., Fischer, P. M., & Rojas, T. H. (1994). Do adolescents attend to warnings in cigarette advertising? An eye-tracking approach. *Journal of Advertising Research*, 34, 39–51.

Kruschke, J. K. (2003). Attention in learning. *Current Directions in Psychological Science*, *12*(5), 171–175.

Leong, C. W., Chen, L., Feng, G., Lee, C. M., & Mulholland, M. (2015). Utilizing depth sensors for analyzing multimodal presentations: Hardware, software and toolkits. *Proceedings of the 17th ACM International Conference on Multimodal Interaction* (ICMI '15), 9–13 November 2015, Seattle, WA, USA (pp. 547–556). New York: ACM.

Lin, Y.-T., Lin, R.-Y., Lin, Y.-C., & Lee, G. C. (2013). Real-time eye-gaze estimation using a low-resolution webcam. *Multimedia Tools and Applications*, *65*(3), 543–568.

Lubold, N., & Pon-Barry, H. (2014). Acoustic-prosodic entrainment and rapport in collaborative learning dialogues. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (MLA '14), 12–16 November 2014, Istanbul, Turkey (pp. 5–12). New York: ACM.

Lund, K. (2007). The importance of gaze and gesture in interactive multimodal explanation. *Language Resources and Evaluation*, *41*(3–4), 289–303.

Luzardo, G., Guamán, B., Chiluiza, K., Castells, J., & Ochoa, X. (2014). Estimation of presentations skills based on slides and audio features. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (MLA '14), 12–16 November 2014, Istanbul, Turkey (pp. 37–44). New York: ACM.

Luz, S. (2013). Automatic identification of experts and performance prediction in the multimodal math data corpus through analysis of speech interaction. *Proceedings of the 15th ACM International Conference on Multimodal Interaction* (ICMI '13), 9–13 December 2013, Sydney, Australia (pp. 575–582). New York: ACM.

Markaki, V., Lund, K., & Sanchez, E. (2015). Design digital epistemic games: A longitudinal multimodal analysis. Paper presented at the conference *Revisiting Participation: Language and Bodies in Interaction*, 24–27 June 2015, Basel, Switzerland.

Mazur-Palandre, A., Colletta, J. M., & Lund, K. (2014). Context sensitive "how" explanation in children's multimodal behavior, *Journal of Multimodal Communication Studies*, 2, 1–17.

Mishra, B., Fernandes, S. L., Abhishek, K., Alva, A., Shetty, C., Ajila, C. V., … Shetty, P. (2015). Facial expression recognition using feature based techniques and model based techniques: A survey. *Proceedings of the 2nd International Conference on Electronics and Communication Systems* (ICECS 2015), 26–27 February 2015, Coimbatore, India (pp. 589–594). IEEE.

Mitra, S., & Acharya, T. (2007). Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(3), 311–324.

Morency, L.-P., Oviatt, S., Scherer, S., Weibel, N., & Worsley, M. (2013). ICMI 2013 grand challenge workshop on multimodal learning analytics. *Proceedings of the 15th ACM International Conference on Multimodal Interaction* (ICMI '13), 9–13 December 2013, Sydney, Australia (pp. 373–378). New York: ACM.

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Ochoa, X., Chiluiza, K., Méndez, G., Luzardo, G., Guamán, B., & Castells, J. (2013). Expertise estimation based on simple multimodal features. *Proceedings of the 15th ACM International Conference on Multimodal Interaction* (ICMI '13), 9–13 December 2013, Sydney, Australia (pp. 583–590). New York: ACM.

Ochoa, X., Worsley, M., Chiluiza, K., & Luz, S. (2014). MLA '14: Third multimodal learning analytics workshop and grand challenges. *Proceedings of the 16th ACM International Conference on Multimodal Interaction* (ICMI '14), 12–16 November 2014, Istanbul, Turkey (pp. 531–532). New York: ACM.

Oviatt, S., Cohen, A., & Weibel, N. (2013). Multimodal learning analytics: Description of math data corpus for ICMI grand challenge workshop. *Proceedings of the 15th ACM International Conference on Multimodal Interaction* (ICMI '13), 9–13 December 2013, Sydney, Australia (pp. 583–590). New York: ACM.

Oviatt, S., & Cohen, P. R. (2015). *The paradigm shift to multimodality in contemporary computer interfaces*. San Rafael, CA: Morgan & Claypool Publishers.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. *Encyclopedia of Human Computer Interaction*, 1, 211–219.

Raca, M., & Dillenbourg, P. (2013). System for assessing classroom attention. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 265–269). New York: ACM.

Raca, M., Tormey, R., & Dillenbourg, P. (2014). Sleepers' lag: Study on motion and attention. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 36–43). New York: ACM.

Scherer, S., Worsley, M., & Morency, L.-P. (2012). 1st international workshop on multimodal learning analytics. *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (ICMI '12), 22–26 October 2012, Santa Monica, CA, USA (pp. 609–610). New York: ACM.

Schlömer, T., Poppinga, B., Henze, N., & Boll, S. (2008). Gesture recognition with a Wii controller. *Proceedings of the 2nd International Conference on Tangible and Embedded Interaction* (TEI '08), 18–21 February 2008, Bonn, Germany (pp. 11–14). New York: ACM.

Schneider, J., Börner, D., van Rosmalen, P., & Specht, M. (2015). Presentation trainer, your public speaking multimodal coach. *Proceedings of the 17th ACM International Conference on Multimodal Interaction* (ICMI '15), 9–13 November 2015, Seattle, WA, USA (pp. 539–546). New York: ACM.

Serrano-Laguna, A., & Fernández-Manjón, B. (2014). Applying learning analytics to simplify serious games deployment in the classroom. *Proceedings of the 2014 IEEE Global Engineering Education Conference* (EDUCON 2014), 3–5 April 2014, Istanbul, Turkey (pp. 872–877). IEEE.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 9, 51–60. doi:0002764213498851.

Silver, E. A. (2013). *Teaching and learning mathematical problem solving: Multiple research perspectives*. Routledge.

Simsek, D., Sándor, Á., Buckingham Shum, S., Ferguson, R., De Liddo, A., & Whitelock, D. (2015). Correlations between automated rhetorical analysis and tutors' grades on student essays. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 355–359). New York: ACM.

Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. *Big Data*, *1*(2), 85–99.

Thompson, K. (2013). Using micro-patterns of speech to predict the correctness of answers to mathematics problems: An exercise in multimodal learning analytics. *Proceedings of the 15th ACM International Conference on Multimodal Interaction* (ICMI '13), 9–13 December 2013, Sydney, Australia (pp. 591–598). New York: ACM.

Vasquez, H. A., Vargas, H. S., & Sucar, L. E. (2015). Using gestures to interact with a service robot using Kinect 2. *Advances in Computer Vision and Pattern Recognition*, 85. Springer.

Vatrapu, R., Reimann, P., Bull, S., & Johnson, M. (2013). An eye-tracking study of notational, informational, and emotional aspects of learning analytics representations. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 125–134). New York: ACM.

Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 94–101). New York: ACM.

Worsley, M., & Blikstein, P. (2014a). Deciphering the practices and affordances of different reasoning strategies through multimodal learning analytics. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (MLA '14), 12–16 November 2014, Istanbul, Turkey (pp. 21–27). New York: ACM.

Worsley, M., & Blikstein, P. (2014b). Using multimodal learning analytics to study learning mechanisms. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 431–432). International Educational Data Mining Society.

Worsley, M., & Blikstein, P. (2015a). Leveraging multimodal learning analytics to differentiate student learning strategies. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 360–367). New York: ACM.

Worsley, M., & Blikstein, P. (2015b). Using learning analytics to study cognitive disequilibrium in a complex learning environment. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 426–427). New York: ACM.

Zhang, Z. (2012). Microsoft Kinect sensor and its effect. *IEEE MultiMedia*, *19*(2), 4–10.

Zhou, J., Hang, K., Oviatt, S., Yu, K., & Chen, F. (2014). Combining empirical and machine learning techniques to predict math expertise using pen signal features. *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge* (MLA '14), 12–16 November 2014, Istanbul, Turkey (pp. 29–36). New York: ACM.

Chapter 12: Learning Analytics Dashboards

Joris Klerkx, Katrien Verbert, and Erik Duval

Department of Computer Science, KU Leuven, Belgium
DOI: 10.18608/hla17.012

ABSTRACT

This chapter presents learning analytics dashboards that visualize learning traces to give users insight into the learning process. Examples are shown of what data these dashboards use, for whom they are intended, what the goal is, and how data can be visualized. In addition, guidelines on how to get started with the development of learning analytics dashboards are presented for practitioners and researchers.

Keywords: Information visualization, learning analytics dashboards

In recent years, many learning analytics dashboards have been deployed to support insight into learning data. The objectives of these dashboards include providing feedback on learning activities, supporting reflection and decision making, increasing engagement and motivation, and reducing dropout. These learning analytics dashboards apply information visualization techniques to help teachers, learners, and other stakeholders explore and understand relevant user traces collected in various (online) environments. The overall objective is to improve (human) learning.

The goal of this chapter is to provide a guide to practitioners and researchers who want to get started with the development and evaluation of learning analytics dashboards. We provide guidance, and several examples, to address the following items:

1. What kind of data can be visualized?
2. For whom are the visualizations intended (learner, teacher, manager, researcher, other)?
3. Why: what is the goal of the visualization?
4. How can the data be visualized? Which interaction techniques can be applied? What tools, libraries, data formats, et cetera can be used for the technical implementations? What workflow and recipe can be used to develop the visualization?

In addition to these four questions, we elaborate on evaluation aspects that assess the usefulness and potential impact of the approach

BACKGROUND

To Augment the Human Intellect

There is a strong contrast between intelligent systems that try to make decisions on behalf of people, such as intelligent tutoring systems (Brusilovsky, 2000) and educational data mining systems (Santos et al., 2015), and systems that try to empower people to make better-informed decisions. For instance, visual analytics systems (Shneiderman & Bederson, 2003) provide a clear overview of the context, the decisions that can be made, and the potential implications of those decisions.

Data mining plays to the strength of computers to do number crunching, while visualization techniques play to the remarkable perceptual abilities that humans possess. The difference between the two approaches is like the difference between a self-driving car and a car with a human driver. Data mining uses automatic pattern matching for remote control while the dashboard provides visual communication to assist a human driver who remains in control of the vehicle.

There is a certain philosophical or ethical side to this notion of two approaches as well: if learners are always told what to do next, how can they develop the typical 21st-century skills of collaboration, communication, critical thinking, and creativity? Or, at a more fundamental level, how can they become citizens equipped with the knowledge, skills, and attitudes to participate fully in society? In this chapter, we focus on methods that augment the human intellect, through visual approaches for learning analytics (Engelbart, 1995).

Information Visualization

Information visualization is the use of interactive visual representations to amplify cognition (Card, Mackinlay, & Shneiderman, 1999). It typically focuses on abstract data without a straightforward representation in 2-D or 3-D space. Visual analytics puts specific emphasis on building models and visualizing these in order to better understand or refine the models. A very useful goal of information visualization is to rely on human perceptual abilities for pattern discovery (trends, gaps, outliers, clusters). These patterns often become more apparent visually than numerically. As Ware (2004) explains it:

> The human visual system is a pattern seeker of enormous power and subtlety. The eye and the visual cortex of the brain form a massively parallel processor that provides the highest-bandwidth channel into human cognitive centers. At higher levels of processing, perception and cognition are closely interrelated, which is the reason why the words "understanding" and "seeing" are synonymous. (p. xvi)

As such, visualization has the potential to be more precise and revealing than conventional statistical computations (Tufte, 2001).

Static visualizations (i.e., an image) typically provide answers to a limited number of questions that a user might have about a data set. For example, so-called infographics are often used for storytelling in journalism. However, looking at an evocative visualization often leads to new questions that can only be answered by interacting with the data itself (Few, 2009). Adding dynamic interaction techniques to the visualization, therefore, is often necessary to design meaningful visualization tools that encourage exploratory data analysis.

Another advantage of visualization is the ability to reveal problems with the data itself; for instance, about the way the data has been collected. Especially in the case of learning analytics, where (semi-) automated trackers often capture traces of learner activities, this advantage is valuable for quality control.

WHAT FOR, WHOM, WHY, HOW?

What follows is a non-exhaustive overview; it is important to recognize the variety of approaches. This variety is not surprising given the wide variety of learning analytics data that can be visualized, for a wide variety of audiences and reasons, in a wide variety of ways.

Verbert et al. (2014) present a survey of learning analytics dashboard applications "ranging from small mobile applications to learnscapes on large public displays" (p. 1499). Dashboards, they say, "typically capture and visualize traces of learning activities, in order to promote awareness, reflection, and sense-making, and to enable learners to define goals and track progress towards these goals" (p. 1499). The paper makes useful distinction between various types of dashboards:

1. Dashboards that support traditional face-to-face lectures, so as to enable the teacher to adapt the teaching, or to engage students during lecture sessions.
2. Dashboards that support face-to-face group work and classroom orchestration, for instance by visualizing activities of both individual learners and groups of learners.
3. Dashboards that support online or blended learning: an early famous example is Course Signals that visualizes predicted learning outcomes as a traffic light, based on grades in the course so far, time on task and past performance (Arnold & Pistilli, 2012).

   More sophisticated and complex visualizations for detailed analysis of course activity by teachers are the focus of the Student Activity Meter (Govaerts, Verbert, Duval, & Pardo, 2012). SNAPP focuses on the visualization of social activity of learners (Bakharia & Dawson, 2011).

In terms of what is being tracked, the possibilities continue to expand, as new online trackers become available, capturing more detail of what learners and teachers do. As well, new sensors proliferate that can likewise capture what people do in the analog world. This second data source is evolving especially rapidly, with mobile devices that now include sensors to report physiological, emotional, and other kinds of learner characteristics that have so far mostly eluded automated capturing. Besides tracking, self-reporting can also be a valuable source of data. Although more error-prone and difficult to sustain systematically, self-reporting offers an opportunity for awareness, reflection, and self-analysis.

As for what can be incorporated into a dashboard, Verbert et al. (2014) lists the following kinds of data:

1. *Artefacts produced* by learners, including blog posts, shared documents, software, and other artefacts that would often end up in a student project portfolio.
2. *Social interaction*, including speech in face-to-face group work, blog comments, Twitter or discussion forum interactions.
3. *Resource use* can include consultation of documents (manuals, web pages, slides), views of videos, et

Figure 12.1. (Top) Navi Badgeboard – Personal Badge Overview: A student's badge overview for a given period; (bottom) Navi Surface: students actively using the tabletop display application during a face-to-face session (Charleer et al., 2013).

cetera. Techniques like software trackers and eye-tracking can provide detailed information about what parts of resources exactly are being used and how.

4. *Time spent* can be useful for teachers to identify students at risk and for students to compare their own efforts with those of their peers.
5. *Test and self-assessment* results can provide an indication of learning progress.

Figure 12.1 presents one of our more recent dashboards (Charleer, Klerkx, Odriozola, Luis, & Duval, 2013). The dashboard tracks social data from blogs and Twitter. Such data, categorized as *artefacts produced*, is then visualized for *students*. The goal is to support *awareness about learning progress* and to *enable discussion* in class. To support such awareness and discussion, social interactions of students are abstracted in the form of learning badges for students to earn. Students can then explore which badges they have earned (Figure 12.1, top) through the visualization of *icons* and *colour cues*. Gray badges have not yet been earned. The bottom part of Figure 12.1 shows a visualization, developed for collaborative use on a tabletop that uses a node *link diagram* to enable further exploration of these badges. Among other things, students can explore which other students have earned specific badges as a means to compare and discuss learning progress.

Figure 12.2 shows a dashboard that uses grades to predict a student's chances of failing a particular course (Ochoa, Verbert, Chiluiza, & Duval, 2016) before she starts. The dashboard is intended to support teachers in giving advice to students on their learning trajectories. More specifically, the dashboard presents the likelihood (68%) of this particular student failing a course in which she is interested. The dashboard uses colour cues to indicate whether the risk of failure, based on past performance, is low (green), medium (yellow), or high (red). Depending on the outcome, the teacher can advise the student to take the course or to discuss alternatives, such as first taking a prerequisite course. The dashboard also supports several interaction techniques that enable the teacher to indicate which data should be taken into account to generate this prediction, including sliders at the bottom that enable the teacher to specify the range of data in terms of years. For example, if a student did poorly in Biology in Grade 10 but worked harder and did well in Grade

Figure 12.2. Muva dashboard that represents the likelihood of failing a specific course (Ochoa et al., 2016).

12, the Grade 10 mark can be disregarded.

HOW TO GET STARTED

To leverage the advanced perceptual abilities of humans to help them explore and discover patterns, a designer must create a visual representation or encoding of the data (Card et al., 1999). Several steps, outlined below, can be distinguished in this design process.

Understand Your Goals

The first step is getting to know the problem domain, the data set, the intended end-users of the tool, the typical tasks they should be able to perform, and so on. The following questions need to be answered at this stage:

1. *Why*: What is the goal of the visualization? What questions about the data should it answer?
2. *For whom*: For whom is the visualization intended? Are the people involved specialists in the domain, or in visualization?
3. *What*: What data will the visualization display? Do these data exhibit a specific internal structure, like time, a hierarchy, or a network?
4. *How*: How will the visualization support the goal? How will people be able to interact with the visualization? What is the intended output device?

By carefully examining and understanding the data set, a variety of questions about the data can be formed. Having these questions in mind can be useful when acquiring and filtering data for the dashboard. For example, consider a data set that contains the following learner traces:

- access to learning resources
- time on page in digital textbooks
- contributions to discussion fora
- time spent on assignments

From these traces, we can define several relevant questions as a starting point in the design process. A teacher might ask questions like these:

- When did students start looking at the course material?
- What is the average time that a student spends reading the textbook?
- How many hours did Peter work on his assignment?
- How often did Peter ask a question on the discussion forum?

A student will probably ask similar questions:

- How much time do I spend on an assignment, compared to other students?
- How much do I contribute to the discussion forum, compared to other students?

In both cases, we deliberately only list questions that start with "what," "when," "how much," and "how often." These specific, direct questions can be directly mapped in a data set. Questions like, "*Why did this student have to enroll twice in this course*?" the answer is more exploratory in nature. Indicators may be that he did not spend enough time on the course material, did not interact with fellow students on the discussion forum, started to study the course material too late, and so on. Another difficult question to answer would be, "*Are students more eager to work on assignment 1 or assignment 2*?" Even if much data is captured, it is difficult to answer questions involving human motivations based on a plurality of (un)known variables. Especially in the early phase of design, it is therefore often advisable and easier to focus on direct, specific questions.

Acquire and (Pre-)Process Your Data

Building a visual dashboard typically entails a data-gathering and preprocessing step. Visualization experts suggest that this step takes 80% of the time and effort versus all other steps. McDonnel and Elmqvist (2009) identify the following intermediary steps:

1. **Acquiring raw data**: It is important to have a clear idea of where the data will come from (e.g., the log files of the LMS, assessment results, other), and when the data will be updated (continuously, not at all, at specific intervals). Will the data be available through an Application Programming Interface (API), an export file, or some other source?
2. **Analyzing raw data**: Data may need to be cleaned if some values are missing or erroneous, or pre-processed to compute aggregate values (mean, minimum, maximum, et cetera). In data analysis, distribution can also be an issue: are there apparent outliers, clusters, et cetera?
3. **Preparing and filtering data**: Using the initial questions from step 1, choose the relevant data from the pool of analyzed raw data.

Mapping Design

Important in the visual mapping design is to choose a representation that best answers the questions you want users to be able to answer, i.e., that serve your visualization goal for the intended target audience. There exists a multitude of alternatives. One way to start is to look at the measurement or scale of each data characteristic. Nominal or qualitative scales differentiate objects based on discrete input domains, such as categories or other qualitative classifications to which they belong. Quantitative scales have continuous input domains (e.g., [0, 100]). Ordinal scales have discrete input domains where the order of the elements matters but the exact difference between the values does not. Depending on the scale of the data characteristic, one can choose how to encode this data visually. Figure 12.3 depicts Mackinlay's (1986) ranking of visual properties to encode quantitative, ordered, and categorical scales. For instance, the spatial position of an element is useful for encoding quantitative, ordered, and categorical differences. This is why scatterplots have been used so often to convey a variety of information. Length, on the other hand, can encode quantitative differences, but is of less value for encoding ordered and categorical differences. Shape is at the bottom of the ranking for visualizing quantitative and ordered differences, but is more often used to depict categorical data.

Low-fidelity prototypes such as paper sketches are often helpful during the design-mapping step. Figure 12.4 depicts an exercise given to the participants of the "Bring Your Own Data: Visual Learning Analytics" tutorial organized at the Learning Analytics Summer Institute (LASI) 2014. Participants included researchers with good knowledge in learning analytics, but limited knowledge about visualization. They were asked to take 15 minutes to sketch all possible ways to visualize a simple data set of two numbers {75, 37}. The exercise illustrated to participants that from the moment they start sketching, it is not difficult to brainstorm visual encodings of data. This is reflected in the number of sketches that two teams of two persons each were able to generate in 15 minutes (see Figure 12.4a and 12.4b).

By sketching, more ideas and questions about the data set are often raised, which in turn leads to new ideas for visualization. For example:

- Figure 12.4c: participants represented the difference between the numbers quite originally by relating them to age, where a person of 37 can easily lift weights, while a person of 73 might already need a walking stick.
- Figure 12.4d: adds muscle size.
- Figure 12.4e: uses shading of an equally sized circle with 75 versus 37 stripes.
- Figure 12.4f: uses a position in a Cartesian coordinate system.
- Figure 12.4g: visualizes a part-to-whole relationship between the numbers.
- Figure 12.4h: assumes a time-based relationship between both numbers, which leads to a negative trend line.
- Figure 12.4i: uses point clouds.
- Figure 12.4j: visualizes an unbalanced scale to represent a difference in weight.
- Figure 12.4k: correlates the size of the figure with the size of the number.

After selecting a visual encoding, high-fidelity prototypes can be built using visualization tools (like Tableau, or even Microsoft Excel) or existing visualization libraries (like Processing or D3.js).

Clearly some alternatives work better than others, depending on the contextualization (e.g., weight and age) and the ability to be interpreted by users (e.g., the mental model of a balanced scale). There is, therefore, no best way to visualize a data set, but some techniques have been proven to work better than others, for example:

- Pie charts are usually a bad idea (Few, 2009).
- Bar charts can be quite powerful.
- Coordinated graphs enable rich exploration.

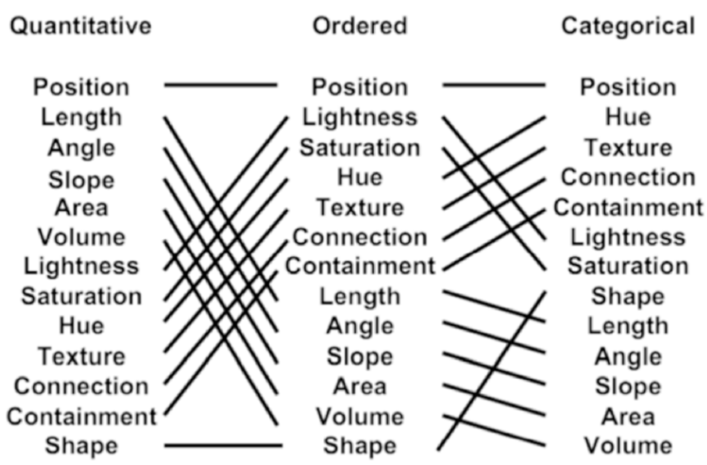


Figure 12.3. Mackinlay's (1986) ranking of visual properties for data characteristics on quantitative, ordered, and categorical scales.

- 3-D graphics often do not convey any additional information and force the reader to deal with redundant and extraneous cues (Levy, Zacks, Tversky, & Schiano, 1996).
- Scatterplots and parallel coordinates are good representations for depicting correlations. In addition, Harrison, Yang, Franconeri, & Chang (2014) found that among the stacked chart variants, the stacked bar significantly outperformed both the stacked area and stacked line. Elliot (2016) has presented a nice overview of these studies.

Documentation

As with any design exercise, it is important to be explicit about:

1. **Rationale:** Why were certain decisions made, what was the intent?
2. **Alternatives:** Which alternatives were considered and why were they not withheld?
3. **Evolution:** How has the design evolved from early sketches to a full-blown implementation? What was modified for conceptual reasons and what for implementation or other reasons (logistics, lack of time, other reasons)?

Add Interaction Techniques

Visual analysis typically progresses in an iterative process of view creation, exploration, and refinement (Heer & Shneiderman, 2012). Before analyzing which interaction techniques are useful for a specific visualization application, it is useful to understand the typical analytical tasks performed by teachers who want to understand how their students are doing in class. Several task taxonomies have been described in literature for this purpose. Common tasks include:

- Comparing values and patterns to find similarities and differences.
- Sorting items based on a variety of data values or metrics.
- Filtering values that satisfy a set of conditions.
- Highlighting data to make specific values stand out visually without making all other data disappear, as is the case with filtering data.
- Clustering or grouping similar items together; for example, by aggregating quantitative data (e.g., average, count, et cetera) to view it in a higher or lower level of detail.
- Annotating findings and thoughts.
- Bookmarking or recording a specific view on the data to enable effective navigation.

Heer and Shneiderman (2012) is essential reading on interactive dynamics for visual analytics. The authors present a taxonomy of interactive dynamics that contribute to successful visual analytic tools. For each task category, various existing visualization systems are described with useful interaction techniques that support the task at hand, such as brushing and linking, histogram sliders, zoomable maps, dynamic query filter widgets, small multiple displays or trellis plots, multiple coordinated views, visual analysis histories, and so on.

Evaluate Continuously

During the design process, the elaboration of concrete personas and scenarios can be very rewarding as it helps to focus the design, development, and evaluation of the visualization on what is relevant. It is very easy to get carried away with too much "eye candy" and lose track of the what, for whom, and why the visualization

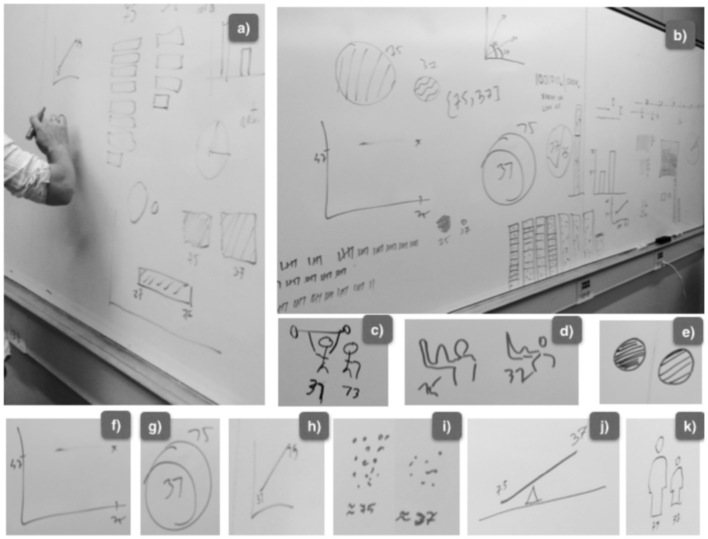


Figure 12.4. Sketches of a small data set of two numbers {75, 37}.

is being designed. Generally, a user-centred design (UCD) approach proceeds with iterative development that keeps the target users in the loop in continuous cycles of design–implementation–evaluation. In this way, the development can focus on the most relevant issues for teachers or learners at all times.

The evaluation of information visualization systems is essential. A plethora of techniques can be used, including controlled experiments that evaluate different visualization and interaction techniques or field studies that assess the impact of a visualization on learning (Plaisant, 2004). The latter take place in natural environments (classrooms) but are often time consuming and difficult to replicate and generalize (Nagel et al., 2014). Verbert et al. (2014) suggest the following evaluation techniques:

1. Effectiveness, which can refer to engagement, higher grades or post-test results, higher retention rates, improved self-assessment, and overall course satisfaction.
2. Efficiency in the use of time of a teacher or learner.
3. Usability and usefulness evaluations often focus on teachers being able to identify learners at risk or asking learners how well they think they are performing in a course.

Typical evaluation instruments include questionnaires or controlled experiments where time-to-task, errors made, time-to-learn, et cetera are evaluated (Dillenbourg et al., 2011).

CONCLUSION

Information visualization concepts and methodologies are key enablers for

- Learners to gain insight into their learning actions and the effects these have.
- Teachers to stay aware of the subtle interactions in their courses.
- Researchers to discover patterns in large data sets of user traces and to communicate these data to their peers.

As shown in this chapter, visualization has the unique potential to help shape the learning process and encourage reflection on its progress and impact by creating learning analytics dashboards that give a concise overview of relevant metrics in an actionable way and that support the exploration of patterns.

Designing and creating an effective information visualization system for learning analytics is an art, as the designer needs both domain expertise on learning theories and paradigms as well as techniques ranging from visual design to algorithm design (Nagel, 2015; Spence, 2001). In this chapter, we have briefly introduced the various steps in a visualization design process, from raw data analysis to effective dashboards evaluated by target users.

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue: Using learning analytics to increase student success. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 267–270). New York: ACM.

Bakharia, A., & Dawson, S. (2011). SNAPP: A bird's-eye view of temporal participant interaction. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 168–173). New York: ACM.

Brusilovsky, P. (2000). Adaptive hypermedia: From intelligent tutoring systems to web-based education. In G. Gauthier, C. Frasson, K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems* (ITS 2000), 19–23 June 2000, Montreal, QC, Canada (pp. 1–7). Springer. doi:10.1007/3-540-45108-0_1

Card, S. K., Mackinlay, J., & Shneiderman, B. (Eds.). (1999). *Readings in information visualization: Using vision to think*. Burlington, MA: Morgan Kaufmann Publishers.

Charleer, S., Klerkx, J., Odriozola, S., Luis, J., & Duval, E. (2013). Improving awareness and reflection through collaborative, interactive visualizations of badges. In M. Kravcik, B. Krogstie, A. Moore, V. Pammer, L. Pannese, M. Prilla, W. Reinhardt, & T. D. Ullmann (Eds.), *Proceedings of the 3rd Workshop on Awareness and Reflection in Technology Enhanced Learning* (AR-TEL '13), 17 September 2013, Paphos, Cyprus (pp. pp. 69–81).

Dillenbourg, P., Zufferey, G., Alavi, H., Jermann, P., Do-lenh, S., Bonnard, Q., Cuendet, S., & Kaplan, F. (2011). Classroom orchestration: The third circle of usability. *Proceedings of the 9th International Conference* *on Computer-Supported Collaborative Learning* (CSCL 2011), 4–8 July 2011, Hong Kong, China (vol. I, pp. 510–517). International Society of the Learning Sciences.

Elliot, K. (2016). 39 studies about human perception in 30 minutes. Presented at OpenVis 2016. https://medium.com/@kennelliott/39-studies-about-human-perception-in-30-minutes-4728f9e31a73#.rs7z3ch6i

Engelbart, D. (1995). Toward augmenting the human intellect and boosting our collective IQ. *Communications of the ACM*, 38(8), 30–34.

Few, S. (2009). *Now you see it: Simple visualization techniques for quantitative analysis*. Berkeley, CA: Analytics Press.

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The student activity meter for awareness and self-reflection. *CHI Conference on Human Factors in Computing Systems: Extended Abstracts* (CHI EA '12), 5–10 May 2012, Austin, TX, USA (pp. 869–884). New York: ACM.

Harrison, L., Yang, F., Franconeri, S., & Chang, R. (2014). Ranking visualizations of correlation using Weber's law. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1943–1952.

Heer, J., & Shneiderman, B. (2012, February). Interactive dynamics for visual analysis. *Queue – Microprocessors*, 10(2), 30.

Levy, E., Zacks, J., Tversky, B., & Schiano, D. (1996, April). Gratuitous graphics? Putting preferences in perspective. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '96), 13–18 April 1996, Vancouver, BC, Canada (pp. 42–49). New York: ACM.

Mackinlay, J. (1986). Automating the design of graphical presentations of relational information. *Transactions on Graphics*, 5(2), 110–141.

McDonnel, B., & Elmqvist, N. (2009). Towards utilizing GPUs in information visualization: A model and implementation of image-space operations. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 1105–1112.

Nagel, T., Maitan, M., Duval, E., Vande Moere, A., Klerkx, J., Kloeckl, K., & Ratti, C. (2014). Touching transport: A case study on visualizing metropolitan public transit on interactive tabletops. *Proceedings of the 12th International Working Conference on Advanced Visual Interfaces* (AVI 2014), 27–29 May 2014, Como, Italy (pp. 281–288). New York: ACM.

Nagel, T. (2015). *Unfolding data: Software and design approaches to support casual exploration of tempo-spatial data on interactive tabletops*. Leuven, Belgium: KU Leuven, Faculty of Engineering Science.

Ochoa, X., Verbert, K. Chiluiza, K., & Duval, E. (2016). Uncertainty visualization in learning analytics. Work in progress for *IEEE Transactions on Learning Technologies*.

Plaisant, C. (2004). The challenge of information visualization evaluation. *Proceedings of the 2nd International Working Conference on Advanced Visual Interfaces* (AVI '04), 25–28 May 2004, Gallipoli, Italy (pp. 109–116). New York: ACM.

Santos, O. C., Boticario, J. G., Romero, C., Pechenizkiy, M., Merceron, A., Mitros, P., Luna, J. M., Mihaescu, C., Moreno, P., Hershkovitz, A., Ventura, S., & Desmarais, M. (Eds.). (2015). *Proceedings of the 8th International Conference on Educational Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain. International Educational Data Mining Society.

Shneiderman, B., & Bederson, B. B. (2003). *The craft of information visualization: Readings and reflections*. Burlington, MA: Morgan Kaufmann Publishers.

Spence, R. (2001). *Information visualization: Design for interaction*, 1st ed. Salt Lake City, UT: Addison-Wesley.

Tufte, E. R. (2001). *The visual display of quantitative information*, 2nd ed. Cheshire, CT: Graphics Press LLC.

Verbert, K., Govaerts, S., Duval, E., Santos, J., Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6), 1499–1514.

Ware, C. (2004). *Information visualization: Perception for design*, 2nd ed. Burlington, MA: Morgan Kaufmann Publishers.

Chapter 13: Learning Analytics Implementation Design

Alyssa Friend Wise, Jovita Vytasek

Learning Analytics Research Network, New York University, USA
Faculty of Education, Simon Fraser University, Canada

DOI: 10.18608/hla17.013

ABSTRACT

This chapter addresses the design of learning analytics implementations: the purposeful shaping of the human processes involved in taking up and using analytic tools, data, and reports as part of an educational endeavor. This is a distinct but equally important set of design choices from those made in the creation of the learning analytics systems themselves. The first part of the chapter reviews key challenges of interpretation and action in analytics use. The three principles of Coordination, Comparison, and Customization are then presented as guides for thinking about the design of learning analytics implementations. The remainder of the chapter reviews the existing research and theory base of learning analytics implementation design for instructors (related to the practices of learning design and orchestration) and students (as part of a reflective and self-regulated learning cycle). Implications for learning analytics designers and researchers and areas requiring further research are highlighted.

Keywords: Learning design, analytics implementation, learning analytics implementation challenges, teachers-facing learning analytics, student-facing learning analytics

Much of the work of learning analytics researchers and designers revolves around the challenges of how to extract, process, and present data in ways that are useful to various educational stakeholders. However, after measures have been created and displays designed, there is still additional work required for analytics to play a constructive role in educational systems. System design alone does not ensure successful uptake (Ertmer, 1999; Hall, 2010; Donnelly, McGarr, & O'Reilly, 2011) as "analytics exist as part of a sociotechnical system where human decision making and consequent actions are as much a part of any successful analytics solution as the technical components" (van Harmelen & Workman, 2012, p. 4). Thus, learning analytics researchers and practitioners need to attend to the human activity of working with these tools and develop a knowledge base for the design of learning analytics implementations (see Figure 13.1).

DEFINING LEARNING ANALYTICS IMPLEMENTATIONS

This chapter focuses on the elements shaping how learning analytics are motivated and mobilized for productive use by instructors, learning designers, and students. The act of introducing learning analytics into an educational environment is called a learning analytics implementation. While the term "learning analytics intervention" has also been used in the past (Lonn, Aguilar, & Teasley, 2015; Wise, 2014), it is a more narrow label that implies learning analytics use as an interruption to regular learning practices that occurs at a specific point in time to address a problem. Implementation is preferred as a more general term that also includes ongoing learning analytics use as a sustained activity incorporated into habitual learning practices (Wise, Vytasek, Hausknecht, & Zhao, 2016). Learning analytics implementation design is then defined globally as the purposeful framing of activity surrounding how analytic tools, data, and reports are

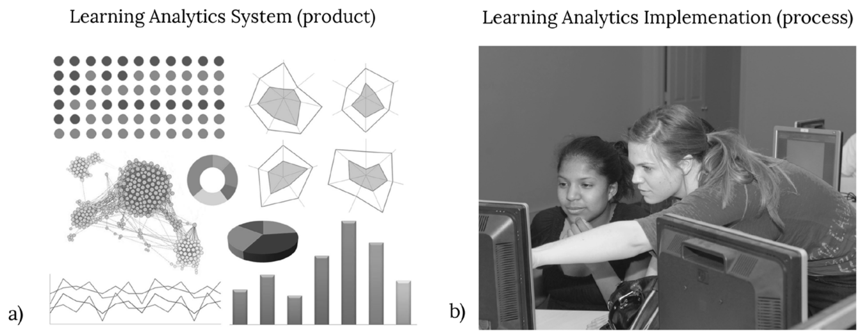


Figure 13.1. Visual differentiation of a) a learning analytics system (product) and b) intentional use of the system by instructors and students (process). Design of the former addresses issues of measures, algorithms, and displays while design of the latter addresses issues of timing, interpretative lens, and action parameters. Source: Photo in b) by US Department of Education licensed under Creative Commons Attribution 2.0 License. Cropped from original (www.flickr.com/photos/departmentofed/9610345404)

taken up and used as part of an educational endeavor. Specifically, it addresses questions of who should have access to particular kinds of analytic data, when the analytics should be consulted, for what purposes, and how the analytics feed back into the larger educational processes taking place.

USING IMPLEMENTATION DESIGN TO ADDRESS LEARNING ANALYTICS CHALLENGES

The process of using learning analytics involves making sense of the information presented and taking action based on it (Siemens, 2013; Clow, 2012). While analytics are often developed for general use across a broad range of situations, the answer to questions of meaning and action are inherently local. Correspondingly, the design of learning analytics implementations needs to be more sensitive to the immediate learning context than the design of learning analytics tools. This is seen in several well-documented challenges in using analytics to inform educational decision-making at the level of *interpretation* as well as at subsequent stages of *taking action* (Wise & Vytasek, in preparation; Wise et al., 2016).

At the level of interpretation, two important challenges are those of context and priorities. *The challenge of context* refers to the fact that analytics are inherently abstracted representations of past activity. Interpreting these representations to inform future activity requires an understanding of the purposes and processes of the learning activity in which they were generated and a mean by which to connect the analytics to these (Lockyer, Heathcote, & Dawson, 2013; Ferguson, 2012). *The challenge of priorities* refers to how users assign relative value to the variety of analytic feedback available. Particular aspects of analytic feedback may be more or less important at different points in the learning process and different analytics can provide information that suggests divergent interpretations that must be reconciled (Wise, 2014).

At the stage of taking action, two important concerns are those of possible options and enacting change. *The challenge of possible options* refers to the fact that analytics provide a retrospective lens to evaluate past activity, but this does not always directly indicate what actions could be taken in the future to change the situation. *The challenge of enacting change* refers to the question of how and on what timeline these actions (once identified) should occur. Change does not occur instantaneously – incremental improvement and intermediate stages of progress need to be considered.

Implementation design helps address these challenges by providing guidance at the mediating level between the analytics presented and the localized course context. This both provides the additional support required to make the information actionable and allows for tailoring of analytics use to meet the needs of particular learning contexts.

IMPLEMENTATION DESIGN CONSIDERATIONS

Learning analytics implementations operate at the interface between the learning activities (the pedagogical events that generate data) and the learning analytics (the designed representations of this data). This relationship can be considered through three guiding principles: Coordination, Comparison, and Customization (Wise & Vytasek, in preparation) grounded in theories of constructivism, metacognition, and self-regulated learning (Duffy & Cunningham, 1996; Schunk & Zimmerman, 2012).

The Principle of Coordination

The principle of Coordination is the foundation of learning analytics implementation design, stating that the surrounding frame of activity through which analytic tools, data, and reports are taken up should position the use of analytics as an integral part of the educational experience tied to goals and expectations (Wise, 2014). To be coordinated with the learning activity, the use of learning analytics needs to be conceived of as a central element of the learning design itself (Lockyer et al., 2013; Pardo, Ellis, & Calvo, 2015; Persico & Pozzi, 2015) so that it is clear to the user how the analytics are meant to play a role in their regular engagement in the learning process.

Conceptual Coordination means an advanced consideration on which of the available analytics to focus (based on the goals of the educational activity) and what productive and unproductive patterns in these metrics are expected to look like (Brooks, Greer, & Gutwin, 2014; Macfadyen & Dawson, 2010; Persico & Pozzi, 2015). To represent the breath of valued actions during a learning activity, it is advisable to use diverse analytic measures (Suthers & Rosen, 2011; Winne & Baker, 2013). It is important to clearly communicate the logic of this connection tying pedagogical goals with learning actions and data-based feedback to the analytics users (Wise, 2014) as initial evidence suggests they put more value on metrics when they clearly understand the connection to learning (Wise, Zhao, & Hausknecht, 2014).

Logistical Coordination means attention to when and how it makes sense for users to work with the chosen analytics as part of the teaching or learning activity. With experienced learning analytics users or those with strong self-monitoring skills, it may be fine to provide only Conceptual Coordination and leave room for individual decisions around when to consult the analytics (van Leeuwen, 2015). However, in many cases, explicit guidance about when and how to work with the analytics as a tool to support learning or teaching is necessary (Koh, Shibani, Tan, & Hong, 2016). General strategies include suggesting a rhythm of analytics use (Wise, Zhao, & Hausknecht, 2013) or a timescale for checkpoints (Lockyer et al., 2013); specific approaches for instructor and student use are discussed in sections 5 and 6.

The Principle of Comparison

The principle of Comparison addresses the need for one or more appropriate reference frames with which to evaluate the meaning of an analytic. For example, the interpretation of a student receiving a particular knowledge assessment (say "25") varies depending on the highest possible score, the performance of the rest of the class, and the level of their prior achievement.

Absolute reference frames for learning analytics provide a fixed standard for comparison that has been set in advance; for example, a set of course expectations (Wise, Zhao, & Hausknecht, 2014). Absolute reference frames can vary in the specificity of the standard set by providing an exact target for a metric or a range of desirable values.

Relative reference frames provide a variable standard that fluctuates over time. One relative reference frame is *peer activity*. This commonly used reference frame sets up comparisons across individuals based on a measure of central tendency or distribution (Corrin & de Barba, 2015; Govaerts, Verbert, Duval, & Pardo, 2012). Another relative reference frame is *parallel activity*, in which comparisons are made across learning events within a course or across courses (Bakharia et al., 2016). In this case, it is critical that the activities being compared are indeed parallel in key ways (e.g., duration, intent, expectations), otherwise the comparisons made may lead to invalid inferences. Finally, a less commonly used but powerful reference frame is *prior activity*, in which comparisons are made for the same individual(s) across time, allowing for the tracking of progress and change (Wise, Zhao, & Hausknecht, 2014).

The Principle of Customization

The principle of Customization emerges from the recognition that there are multiple, disparate, and equally valid needs and paths (and potentially endpoints) for different learning analytics users. Customization of learning analytics to meet these different needs can be thought of in two ways. The first approach is computationally driven and can be thought of as *adaptive learning analytics* (cf. Brusilovsky & Peylo, 2003). As this relates to the design of the learning analytics system rather than the learning analytics implementation, it is not addressed further here. A second approach to personalization is user-driven and can be thought of as *adaptable learning analytics* (cf. Brooks et al., 2014). In this case, the analytics interface allows for different kinds of uses by different individuals who determine themselves which analytics they will attend to and in what way. There is a danger, however, that users may be overwhelmed by the multitude of possible options without a clear basis on which to make choices. Thus implementation design needs to support user agency actively by guiding them in the process of effectively making decisions about how to use the learning analytics provided to meet their own needs and context.

LEARNING ANALYTICS IMPLICATION DESIGN FOR INSTRUCTORS

Instructors are a natural audience for learning analytics as they are often already engaged informally in the activity of examining student learning to inform their practice. Such teacher-inquiry has traditionally depended on qualitative methods of reflection using journals, interviews, peer-observation, student observations and examination of learning artifacts (Lytle & Cochran-Smith, 1990), though interest in the use of student data as evidence to inform this process has been increasing (Wasson, Hanson, & Mor, 2016). Analytics can support instructors in approaching this reflective cycle with more detail, using data as an aid in assessing the impact of teaching decisions on learning activity (Mor, Ferguson, & Wasson, 2015). Different bodies of literature have explored specific process of instructor use of analytics in relation to the practices of learning design, orchestration, and assessment.

One perspective focuses on the use of analytics to inform learning design. From this perspective, instructors document their pedagogical intentions through the design, which then provides the conceptual frame for asking questions and making sense of the information provided by the analytics (Dawson, Bakharia, Lockyer, & Heathcote, 2011). This can facilitate an understanding of the effects of a learning design (or specific instructional approach) on student activity and learning (Dietz-Uhler & Hurn, 2013), which can then feed back into improving the design (Persico & Pozzi, 2015; Mor et al., 2015). The process is a cyclical one in which the analytics make the learning processes undertaken by students visible (Martínez-Monés, Harrer, & Dimitriadis, 2011). A specific model for aligning learning analytics use with learning design was developed by Lockyer et al. (2013) who described how instructors can initially map the learning process supported by their design, pre-identify activity patterns that indicate successful (or unsuccessful) student engagement in the pedagogical design, and then use analytics to track learner progression towards the desired outcomes. An initial example of this cycle in action is given in Brooks et al. (2014) who look at instructors' modifications of their discussion forum practices based on sociograms created from students' speaking and listening activity. A similar cycle as engaged in by course designers of a MOOC is given in Roll, Harris, Paulin, MacFadyen, and Ni (2016). Lockyer et al.'s (2013) model represents a strong application of the principle of Coordination as it makes clear how the use of the analytics is integrally tied to the goals and expectations for learning. It also suggests ways analytics can be worked into an instructor's activity flow, for example by setting up checkpoints. The principle of Comparison is also attended to in the sense that the pre-identified activity patterns serve as an absolute reference frame to gauge progress towards a desired state. Additional comparisons – for example, setting incremental stages to target along the way or using prior activity to judge progress – could also be considered. As this use of learning analytics is directed globally at the effects of a learning design, attention to the principle of Customization is currently limited. However, thinking about how different learning designs might work differently and be more or less effective for different kinds of learners and learning contexts is an exciting area for future consideration.

An alternative conceptualization of instructors' analytics use shifts the focus from looking at data patterns for the course as a whole to looking at differences between students or student groups. From this perspective, the analytics are used in (relatively) real-time as a tool to monitor activity, support the diagnoses of situations needing attention, and prompt instructors to intervene when necessary. This can be thought of as a form of orchestration (Rodríguez-Triana, Martínez-Monés, Asensio-Pérez, & Dimitriadis, 2015) in which instructors use analytics to support their awareness of student activity and adapt their teaching to meet student needs (Feldon, 2007). To address the inherent challenges in doing this (Dyckhoff, Lukarov, Muslim, Chatti, & Schroeder, 2013), van Leeuwen (2015) proposes a two-part model of how instructors can work with analytics in this capacity. First, instructors use the analytics to monitor student activity, specifically noticing important differences across individuals or groups. This is supported by the capabilities of analytics to aggregate information for manageable presentation. Second, instructors use the information to inform their diagnosis of situations, individuals, or groups requiring attention. Working in the context of a learning analytics application for a computer-supported collaborative learning context, van Leeuwen (2015) found initial evidence to support the hypothesis that analytics use would increase both the specificity of instructor diagnoses and inform the actions that they took. This model represents a strong application of the principle of Customization as the goal of instructors' analytics use is individualized actions tailored to particular student or group needs. With respect to Comparison, in the original conceptualization there is strong reliance on the relative frame of peer activity, though the prior activity of a group or individual are also taken into account. The addition of an absolute standard with which to compare activity could also be considered. An area for future development is the Coordination of this kind of analytics use with the larger purpose and flow of the collaborative learning activity.

A final model for instructor use of learning analytics that has yet to be fully developed is as a tool for assessment. While there is a need for caution in such applications, there are exciting possibilities for using temporal analytics (which capture time-based characteristics of trace data) to move towards a new paradigm of assessment that replaces current point-in-time evaluations of learning *states* with dynamic evaluations of learning *progress* (Molenaar & Wise, 2016). Such an approach is grounded in Comparison with prior activity and offers opportunities for instructors to respond to individual and evolving learning needs. Using analytics collected during the normal course of learning processes to evaluate the development of student understanding in situ also presents an attractive opportunity for assessment that can both meet summative needs and serve formative purposes. However, the conceptual and logistical Coordination of how learning analytics are used for such assessment purposes is critical for adoption, given the importance of decisions often attached to assessment activities.

LEARNING ANALYTICS IMPLICATION DESIGN FOR STUDENTS

Students are an important audience for analytics use for several reasons. First, since student learning is the ultimate goal of educational systems, much of the data collected in learning analytics systems is information generated by or about students. From an ethical perspective, students have the right (and perhaps the responsibility) to review their own data (Pardo & Siemens, 2014; Slade & Prinsloo, 2013). Second, similar to instructors, students are also at the "front line" of learning and thus potentially well-equipped to bring local context to bear in interpreting analytics, as well as make immediate adjustments to their learning processes based on them. Different from instructors, however, students must negotiate between course-wide goals for learning and analytics use and their own personal objectives (Wise, 2014). This explicitly allows for Customization as it adds an additional personalized reference frame for Comparison.

Student use of analytics has been conceptualized primarily in terms of a reflective cycle in which students use their own analytic data to inform their individual learning processes. Drawing on the theories of Schön (1983) and Kolb (1984), Clow (2012) has put forward the general idea of analytics use as an element of reflective practice in which the information provides feedback that students can use to adjust or experiment with changes in their learning activities. The notion of students using analytics to act as "little experimenters" has also been discussed within the self-regulated learning literature (Winne, in press). Drawing on theories of metacognition, this field has a long history of studying and supporting the ways students monitor and take action on their learning as part of a self-regulative process (Zimmerman & Schunk, 2011; Schunk 2008; Boekaerts, Pintrich, & Zeidner, 2000). Students who adopt positive SRL strategies tend to have richer learning interactions and perform better in their studies (Zimmerman, 2008; Pintrich, 2004; Pardo, Han, & Ellis, 2016). While such efforts have traditionally been limited by the challenges and inaccuracies of human memory and recall (Winne, 2010; Azevedo, Moos, Johnson, & Chauncey, 2010), analytics offer the exciting potential to mirror a learner's activity back to them with greater ease and accuracy (Winne & Baker, 2013). From this perspective, learning analytics are conceived of as a way to cue students to effectively monitor and take action on their learning (Roll & Winne, 2015).

Expanding on these ideas, a more specific vision of student learning analytics use has been put forth by Wise and colleagues (Wise et al., 2016; Wise, 2014; Wise, Zhao, & Hausknecht, 2013; 2014). Their Student Tuning Model describes student's learning-analytics-informed reflective practice as grounded in the relationship between the learning activities and the learning analytics. Students work with this relationship continually as they engage in cycles of goal setting, action, reflection, and adjustment. To support this descriptive cycle of analytics use, Wise et al. (2016) have proposed and presented initial validation evidence for a pedagogical framework for designing learning analytics implementations for students. The Align Design framework utilizes elements of Coordination, Comparison, and Customization as described above with an emphasis on the interplay between agency and dialogue with the situation.

In addition to these overarching models of students' learning analytics use, there are other ongoing research efforts proposing targeted pedagogical frameworks for specific learning contexts and exploring particular aspects of how to design learning analytics implementations for students. Koh et al. (2016) have developed the Team and Self-Diagnostic Learning framework for analytics use in the context of collaborative inquiry with secondary students. This framework provides strong process-based Coordination by integrating instructor-guided use of teamwork competency analytics into students' experiential learning cycles. Attention to Comparison takes the form of contrasts of similarities and difference between self- and peer-ratings on six dimensions of teamwork.

Separately, Aguilar (2015) is conducting research at the intersection of the Customization and Comparison principles, examining whether students' mastery or performance orientation to learning can help determine when peer activity is a useful reference frame for evaluating learning analytics. Similarly, research into individual differences has shown that particular goal-orientations are associated with the use of different kinds of self-regulatory strategies generally (Shirazi, Gašević, & Hatala, 2015) and can specifically influence the interpretation and use of different learning analytics visualizations (Beheshitha, Hatala, Gašević, & Joksimović, 2016).

Such findings can have implications for system-driven adaptive analytics in terms of what measures with what references points are helpful (and ethical) to show to particular learners. For example, while the peer reference frame can be can be motivating in showing a student where they stand in relation to others in the class for some students (Beheshitha et al., 2016; Govaerts, et al., 2012), it can be distracting for others (Corrin & de Barba, 2015). Some students find it demotivating to find out they are doing substantially worse than their classmates (Wise, Hausknecht, & Zhao, 2014). Especially for students who are struggling, the ability to document improvement in comparison to their own prior activity may be more powerful than comparison to a distal class mean. In addition, there are questions of which portion(s) of a peer group are most appropriate for comparison in a given situation; for example, should students be shown data for the whole course, only students who are similar to them in some way, or the "top performers" (Beheshitha et al., 2016). The answer will depend on the kind of activity, relevant student characteristics, and the objective for analytics use.

Other researchers are probing more deeply into ways in which learning analytics implementations can be designed to support student Customization in terms of adaptable analytics implementations. For example, Santos, Govaerts, Verbert, and Duval (2012) describe a process in which students articulate individual goals and then track their progress. Ferguson, Buckingham Shum, and Deakin Crick (2011) have used blogs as a tool for creating individually owned reflective spaces in which students can work through the sense-making of the analytics. The need for students to have time to "digest" the meaning of the analytics before taking action is also supported by the findings of Koh et al. (2016), suggesting that appropriate pacing may be a critical aspect in the Coordination of reflective learning analytics use with the overarching learning activities. In contrast, Holman et al. (2015) found that for predictive analytics use focused on course progress, students tended to use the tools to make plans (and follow-through on these plans) mostly in short bursts just prior to major course deadlines.

While the models and research discussed above have primarily conceptualized student learning analytics use as an individual endeavor, there are also intriguing opportunities for students to work with analytics collectively. This follows the tradition of "group awareness" tools, which have been used to facilitate computer-supported collaborative work and learning (Buder, 2011; Janssen & Bodemer, 2013). In this case, the individuals in a group and the group collectively work with analytics to improve their joint learning process through socially shared regulation (Järvelä et al., 2015).

IMPLICATIONS FOR LEARNING ANALYTICS DESIGNERS AND RESEARCHERS

The above discussion has described three principles for designing learning analytics implementations and has presented current research and models of learning analytics use by instructors and students. This framework can also be used to discuss implications for learning analytics design and research more generally.

First, from a systems design perspective, we can anticipate and create features to support implementation possibilities. For example, a tool that allows instructors to associate particular analytics and course goals (and annotate these connections with examples of productive or unproductive patterns) would support the principle of Coordination. Similarly, creating tools that help students track and reflect on the changes in their analytics over time (for example by being able to adjust the time window of the analytics for both current and historical data) could support the principles of Customization. This latter point is of particular importance given the usefulness of prior activity as a reference frame for evaluating progress, but the predominance of analytic dashboards that only provide point-in-time "snapshots."

Second, from a research perspective, in addition to continued work to develop useful analytics systems, inquiry is also needed into how activity using such analytics is best motivated and mobilized, and the factors influencing this process. Practically, this suggests that laboratory studies, which ask people to perform specific tasks or determine particular information with learning analytics tools, can only contribute so much to predicting how instructors and students will work with analytics "in the wild." Thus field-testing new analytics in real educational contexts early on may prove particularly important in developing learning analytics systems and implementations that truly impact teaching and learning. One valuable approach to consider is Design-Based Intervention Research (Penuel, Fishman, Cheng, & Sabelli, 2011), which emphasizes multiple iterations of testing and (re)design of learning innovations in partnership with practitioners to support on-the-ground use and sustainability.

Finally, it is critical to consider the use of analytics as a radically new technology for instructors and students. Careful planning of how the analytics will be introduced, with appropriate up-front guidance, ongoing support, diverse examples, and time for instructors and students to figure out how to integrate this new form of feedback into their practice is needed to translate the promise of learning analytics into reality. Widespread adoption of learning analytics will not occur spontaneously, but initial reports from projects that have used implementation design to educate users and nurture their analytics use are very promising (Koh et al., 2016; Wise et al., 2016).

CONCLUSION

This chapter reflects the current state of the art of learning analytics implementation design. The principles of Coordination, Comparison, and Customization provide a lens to examine the different dimensions of design choice that can affect how analytics feedback is taken up and acted on in particular educational contexts. For instructors, models have been proposed for analytics use to examine and adjust course-wide learning designs as well as to investigate and respond to individual student activity patterns via orchestration. The use of learning analytics for assessment is a potentially exciting but undeveloped area for application. Student use currently takes the form of a reflective, self-regulative cycle, with attention given to particular ways to support this process. Further research into the impact of students' and instructors' individual differences on analytics use and the generation of designs to support their particular needs is a promising area for future work. The final message of this chapter is to emphasize that intentional implementation design is essential, not optional, for learning analytics adoption. If we wish to avoid the fate of too many promising technologies that never made a real impact on education, research into the interplay of human and technological elements influencing analytics use is a critical area for attention in the field moving forward.

REFERENCES

Aguilar, S. (2015). Exploring and measuring students' sense-making practices around representations of their academic information. Doctoral Consortium Poster presented at the 5th International Conference on Learning Analytics and Knowledge (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA. New York: ACM.

Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist*, 45(4), 210–223.

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, R., Williams, D., Dawson, S., & Lockyer, L. (2016). A conceptual framework linking learning design with learning analytics. *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 329–338). New York: ACM.

Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). The role of achievement goal orientations when studying effect of learning analytics visualizations. *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 54–63). New York: ACM.

Boekaerts, M., Pintrich, P., & Zeidner, M. (2000). *Handbook of self-regulation*. San Diego, CA: Academic Press.

Brooks, C., Greer, J., & Gutwin, C. (2014). The data-assisted approach to building intelligent technology enhanced learning environments. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 123–156). New York: Springer.

Brusilovsky, P., & Peylo, C. (2003). Adaptive and intelligent web-based educational systems. *International Journal of Artificial Intelligence in Education*, 13(2), 159–172.

Buder, J. (2011). Group awareness tools for learning: Current and future directions. *Computers in Human Behavior*, 27(3), 1114–1117.

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics & Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 134–138). New York: ACM.

Corrin, L., & de Barba, P. (2015). How do students interpret feedback delivered via dashboards? *Proceedings of the 5th International Conference on Learning Analytics & Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 430–431). New York: ACM.

Dawson, S., Bakharia, A., Lockyer, L., & Heathcote, E. (2011). "Seeing" the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology, 41*(5), 736–752.

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning, 12*(1), 17–26.

Donnelly, D., McGarr, O., & O'Reilly, J. (2011). A framework for teachers' integration of ICT into their classroom practice. *Computers & Education*, 57(2), 1469–1483.

Duffy, T. M., & Cunningham, D. J. (1996). Constructivism: Implications for the design and delivery of instruction. In D. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 170–198). New York: Macmillan.

Dyckhoff, A. L., Lukarov, V., Muslim, A., Chatti, M. A., & Schroeder, U. (2013). Supporting action research with learning analytics. *Proceedings of the 3rd International Conference on Learning Analytics & Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 220–229). New York: ACM.

Ertmer, P. A. (1999). Addressing first- and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development*, 47(4), 47–61.

Feldon, D. F. (2007). Cognitive load and classroom teaching: The double-edged sword of automaticity. *Educational Psychologist, 42*(3), 123–137.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(18), 304–317.

Ferguson, R., Buckingham Shum, S., & Deakin Crick, R. (2011). EnquiryBlogger: Using widgets to support awareness and reflection in a PLE Setting. In W. Reinhardt & T. D. Ullmann (Eds.), *Proceedings of the 1st Workshop on Awareness and Reflection in Personal Learning Environments* (pp. 11–13) Southampton, UK: PLE Conference.

Govaerts, S., Verbert, K., Duval, E., & Pardo, A. (2012). The student activity meter for awareness and self-reflection. *Proceedings of the CHI Conference on Human Factors in Computing Systems, Extended Abstracts* (CHI EA '12), 5–10 May 2012, Austin, TX, USA (pp. 869–884). New York: ACM.

Hall, G. E. (2010). Technology's Achilles Heel: Achieving high-quality implementation. *Journal of Research on Technology in Education, 42*(3), 231–253.

Holman, C., Aguilar, S. J., Levick, A., Stern, J., Plummer, B., & Fishman, B. (2015). Planning for success: How students use a grade prediction tool to win their classes. *Proceedings of the 5th International Conference on Learning Analytics & Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 260–264). New York: ACM.

Janssen, J., & Bodemer, D. (2013). Coordinated computer-supported collaborative learning: Awareness and awareness tools. *Educational Psychologist*, 48(1), 40–55.

Järvelä, S., Kirschner, P. A., Panadero, E., Malmberg, J., Phielix, C., Jaspers, J., & Järvenoja, H. (2015). Enhancing socially shared regulation in collaborative learning groups: Designing for CSCL regulation tools. *Educational Technology Research and Development*, 63(1), 125–142.

Koh, E., Shibani, A., Tan, J. P. L., & Hong, H. (2016). A pedagogical framework for learning analytics in collaborative inquiry tasks: An example from a teamwork competency awareness program. *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 74–83). New York: ACM.

Kolb, D. A. (1984). *Experiential education: Experience as the source of learning and development.* Upper Saddle River, NJ: Prentice Hall.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, 57(10), 1439–1459.

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior*, 47, 90–97.

Lytle, S., & Cochran-Smith, M. (1990). Learning from teacher research: A working typology. *The Teachers College Record*, 92(1), 83–103.

Martínez-Monés, A., Harrer, A., & Dimitriadis, Y. (2011). An interaction-aware design process for the integration of interaction analysis into mainstream CSCL practices. In S. Puntambekar, G. Erkens, & C. Hmelo-Silver (Eds.), *Analyzing interactions in CSCL: Methods, approaches and issues* (pp. 269–291). New York: Springer.

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education*, 54(2), 588–599.

Molenaar, I., & Wise, A. (2016). Grand challenge problem 12: Assessing student learning through continuous collection and interpretation of temporal performance data. In J. Ederle, K. Lund, P. Tchounikine, & F. Fischer (Eds.), *Grand challenge problems in technology-enhanced learning II: MOOCs and beyond* (pp. 59–61). New York: Springer.

Mor, Y., Ferguson, R., & Wasson, B. (2015). Learning design, teacher inquiry into student learning and learning analytics: A call for action. *British Journal of Educational Technology*, 46(2), 221–229.

Pardo, A., Ellis, R. A., & Calvo, R. A. (2015). Combining observational and experiential data to inform the redesign of learning activities. *Proceedings of the 5th International Conference on Learning Analytics & Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 305–309). New York: ACM.

Pardo, A., Han, F., & Ellis, R. A. (2016). Exploring the relation between self-regulation, online activities, and academic performance: A case study. *Proceedings of the 6th International Conference on Learning Analytics & Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 422–429). New York: ACM.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450.

Persico, D., & Pozzi, F. (2015). Informing learning design with learning analytics to improve teacher inquiry. *British Journal of Educational Technology*, 46(2), 230–248.

Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation and design. *Educational Researcher*, 40(7), 331–337.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, 16(4), 385–407.

Rodríguez-Triana, M. J., Martínez-Monés, A., Asensio-Pérez, J. I., & Dimitriadis, Y. (2015). Scripting and monitoring meet each other: Aligning learning analytics and learning design to support teachers in orchestrating CSCL situations. *British Journal of Educational Technology*, 46(2), 330–343.

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, 2(1), 7–12.

Roll, I., Harris, S., Paulin, D., MacFadyen, L. P., & Ni, P. (2016, October). Questions, not answers: Doubling student participation in MOOC forums. Paper presented at *Learning with MOOCs III*, 6–7 October 2016, Philadelphia, PA. http://www.learningwithmoocs2016.org/

Santos, J. L., Govaerts, S., Verbert, K., & Duval, E. (2012). Goal-oriented visualizations of activity tracking: A case study with engineering students. *Proceedings of the 2nd International Conference on Learning Analytics & Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 143–152). New York: ACM.

Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.

Schunk, D. H. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review*, 20(4), 463–467.

Schunk, D. H., & Zimmerman, B. J. (2012). Self-regulation and learning. In I. B. Weiner, W. M. Reynolds, & G. E. Miller (Eds.), *Handbook of psychology*, Vol. 7, *Educational psychology*, 2nd ed. (pp. 45–68). Hoboken, NJ: Wiley.

Shirazi, S., Gašević, D., & Hatala, M. (2015). A process mining approach to linking the study of aptitude and event facets of self-regulated learning. *Proceedings of the 5th International Conference on Learning Analytics & Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 265–269). New York: ACM.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, 57, 1380–1400.

Slade, S., & Prinsloo, P. (2013). Learning analytics ethical issues and dilemmas. *American Behavioral Scientist*, 57(10), 1510–1529.

Suthers, D. D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. *Proceedings of the 1st International Conference on Learning Analytics & Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 64–74). New York: ACM.

van Harmelen, M., & Workman, D. (2012). *Analytics for learning and teaching*. Technical report. CETIS Analytics Series Vol 1. No. 3. Bolton, UK: CETIS.

van Leeuwen, A. (2015). Learning analytics to support teachers during synchronous CSCL: Balancing between overview and overload. *Journal of Learning Analytics*, 2(2), 138–162.

Wasson, B., Hanson, C., & Mor, Y. (2016). Grand challenge problem 11: Empowering teachers with student data. In J. Ederle, K. Lund, P. Tchounikine, & F. Fischer (Eds.), *Grand challenge problems in technology-enhanced learning II: MOOCs and beyond* (pp. 55–58). New York: Springer.

Winne, P. H. (2010). Bootstrapping learner's self-regulated learning. *Psychological Test and Assessment Modeling*, 52(4), 472–490.

Winne, P. H. (in press). Leveraging big data to help each learner upgrade learning and accelerate learning science. *Teachers College Record*.

Winne, P. H., & Baker, R. S. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, 5(1), 1–8.

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *Proceedings of the 4th International Conference on Learning Analytics & Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 203–211). New York: ACM.

Wise, A. F., Hausknecht, S. N., & Zhao, Y. (2014). Attending to others' posts in asynchronous discussions: Learners' online "listening" and its relationship to speaking. International Journal of Computer-Supported Collaborative Learning, 9(2), 185–209.

Wise, A. F., Vytasek, J. M., Hausknecht, S., & Zhao, Y. (2016). Developing learning analytics design knowledge in the "middle space": The student tuning model and align design framework for learning analytics use. *Online Learning Journal* 20(2).

Wise, A. F., & Vytasek, J. M. (in preparation). The three Cs of learning analytics implementation design: Coordination, Comparison, and Customization.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2013). Learning analytics for online discussions: A pedagogical model for intervention with embedded and extracted analytics. *Proceedings of the 3rd International Conference on Learning Analytics & Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 48–56). New York: ACM.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2014). Learning analytics for online discussions: Embedded and extracted approaches. *Journal of Learning Analytics*, 1(2), 48–71.

Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. American Educational Research Journal, 45(1), 166–183.

Zimmerman, B. J., & Schunk, D. H. (Eds.) (2011). *Handbook of self-regulation of learning and performance*. New York: Routledge.

SECTION 3

APPLICATIONS

Chapter 14: Provision of Data-Driven Student Feedback in LA & EDM

Abelardo Pardo,[1] Oleksandra Poquet,[2] Roberto Martínez-Maldonado,[3] Shane Dawson[2]

[1] Faculty of Engineering and IT, The University of Sydney, Australia
[2] Teaching Innovation Unit, University of South Australia, Australia
[3] Connected Intelligence Centre, University of Technology Sydney, Australia

DOI: 10.18608/hla17.014

ABSTRACT

The areas of learning analytics (LA) and educational data mining (EDM) explore the use of data to increase insight about learning environments and improve the overall quality of experience for students. The focus of both disciplines covers a wide spectrum related to instructional design, tutoring, student engagement, student success, emotional well-being, and so on. This chapter focuses on the potential of combining the knowledge from these disciplines with the existing body of research about the provision of feedback to students. Feedback has been identified as one of the factors that can provide substantial improvement in a learning scenario. Although there is a solid body of work characterizing feedback, the combination with the ubiquitous presence of data about learners offers fertile ground to explore new data-driven student support actions.

Keywords: Actionable knowledge, feedback, interventions, student support actions

Over the past two decades, education practice has significantly changed on numerous fronts. This includes shifts in educational policy, the emergence of technology-rich learning spaces, advances in learning theory, and the implementation of quality assurance and assessment, to name but a few. These changes have all influenced how contemporary teaching practice is now enacted and embodied. Despite numerous paradigm shifts in the education space, the key role of feedback in promoting student learning has remained essential to what is viewed as effective teaching. Moreover, with the massification of education, the need for providing real-time feedback and actionable insights to both teachers and learners is becoming increasingly acute. As education embraces digital technologies, there is a widespread assumption that the incorporation of such technologies will further aid and promote student learning and address sociocultural and economic inequities. This positivist ideal reflects the notion that technologies can be adopted to enhance accessibility to education while creating more personalized and adaptive learning pathways.

In this vein, the fields of learning analytics (LA) and educational data mining (EDM) have direct relevance for education. LA and EDM aim to better understand learning processes in order to develop more effective teaching practices (Baker & Siemens, 2014). The analysis of data evolving from student interactions with various technologies to provide feedback on the learner's progression has been central to LA and EDM work. In this chapter, we argue that feedback is one of the most powerful drivers influencing student learning. As such, the overall quality of the learning experience is deeply entwined with the relevance and salience of the feedback a student receives. Moreover, the provision of feedback is closely related to other aspects of a learning experience, such as assessment approaches (Boud, 2000), the learning design (Lockyer, Heathcote, & Dawson, 2013), or strategies to promote student self-regulation (Winne, 2014; Winne & Baker, 2013). Although the majority of the discussion in this chapter can be applied across all educational domains, the review focuses predominantly on post-secondary education and professional development.

THE ROLE OF DATA-DRIVEN FEEDBACK IN LEARNING

Discussions about feedback frequently take place within a framing of assessment and student achievement (Black & Wiliam, 1998; Boud, 2000). In this context, the primary role of feedback is to help the student address any perceived deficits as identified through the completion of an assessment item. Ironically, assessment scores and student achievement data have also become tools for driving political priorities and agendas, and are also used as indicators in quality assurance requirements. Assessment in essence is a two-edged sword used to foster learning as well as a tool for measuring quality assurance and establishing competitive rankings (Wiliam, Lee, Harrison, & Black, 2004). While acknowledging the importance of assessment for quality assurance, we focus specifically on the value of feedback often associated with formative assessment or simply as a component of student completion of set learning tasks. Thus, this chapter explores how student trace data can be exploited to facilitate the transformation of the essence of assessment practices by focusing on feedback mechanisms. With such a purpose, we highlight and discuss current approaches to the creation and delivery of data-enhanced feedback as exemplified through the vast body of research in learning analytics and educational data mining (LA/EDM).

Theoretical Models of Feedback

Although there is no unified definition of feedback in educational contexts, several comprehensive analyses of its effects on learning have been undertaken (e.g., Evans, 2013; Hattie & Timperley, 2007; Kluger & DeNisi, 1996). In sum, strong empirical evidence indicates that feedback is one of the most powerful factors influencing student learning (Hattie, 2008). The majority of studies have concluded that the provision of feedback has positive impact on academic performance. However, the overall effect size varies and, in certain cases, a negative impact has been noted. For instance, a meta-analysis by Kluger and DeNisi (1996) demonstrated that poorly applied feedback, characterized by an inadequate level of detail or the lack of relevance of the provided information, could have a negative effect on student performance. In this case, the authors distinguished between three levels of the locus of learner's attention in feedback: the task, the motivation, and the meta-task level. All three are equally important and can vary gradually in focus. Additionally, Shute (2008) classified feedback in relation to its complexity, and analyzed factors affecting the provision of feedback such as its potential for negative impact, the connection with goal orientation, motivation, the presence in scaffolding mechanisms, timing, or different learner achievement levels. Shute noted that to maximize impact, any feedback provided in response to a learner's action should be non-evaluative, supportive, timely, and specific.

Early models relating feedback to learning largely aimed to identify the types of information provided to the student. Essentially, these studies sought to characterize the effect that different types of information can play on student learning (Kulhavy & Stock, 1989). Initial conceptualizations of feedback were driven by the differences in learning science theorizations of how the gap between the actual and desired state of the learner can be bridged (cf. historical review Kluger & DeNisi, 1996; Mory, 2004). According to Mory (2004), contemporary models build upon pre-existing paradigms by viewing feedback in the context of self-regulated learning (SRL), i.e., *a style of engaging with tasks in which students exercise a suite of powerful skills* (Butler & Winne, 1995). These skills, setting goals, thinking about strategies, selecting the right strategies, and monitoring the effects of these strategies on the progress towards the goals are all associated with student achievement (Butler & Winne, 1995; Pintrich, 1999; Zimmerman, 1990). As part of their theoretical synthesis between feedback and self-regulated learning, Butler and Winne (1995, p. 248) embedded two feedback loops into their model. The first loop is contained within the so-called *cognitive system* and refers to the capacity of individuals to monitor their internal knowledge and beliefs, goals, tactics, and strategies and change them as required by the learning scenario. The second loop occurs when the product resulting from a student engaging with a task is measured, prompting the creation of *external feedback* relayed back to the student; for example, an assessment score, or an instructor commenting upon the completion of a task.

Hattie and Timperley (2007) have provided one of the most influential studies on feedback and its impact on achievement. The authors' conceptual analysis was underpinned by a definition of feedback as *the information provided by an agent regarding the performance or understanding of a student*. The authors proposed a model of feedback articulated around the concept that any feedback should aim to reduce the discrepancy between a student's current understanding and their desired learning goal. As such, feedback can be framed around three questions: *where am I going, how am I going, and where to next*? Hattie and Timperley (2007) proposed that each of these questions should be applied to four different levels: learning task, learning process, self-regulation, and *self*. The learning task level refers to the elements of a simple task; for example, notifying the student if an answer is correct or incorrect. The learning process refers to general learning objectives, including various tasks at different times. The self-regulation level refers to the capacity of reflecting on the learning goals, choosing the right strategy, and monitoring the progress towards those goals. Finally, the *self* level refers to abstract personality traits that may not be related to the learning experience. The process and regulation levels are argued to be the most effective in terms of promoting deep learning and mastery of tasks. Feedback at the task-level is effective only as a supplement to the previous two levels; feedback at the self-level has been shown to be the least effective. These three questions and four levels of feedback provide the right setting to connect feedback with other aspects such as timing, positive vs. negative messages (also referred to as polarity), and the consequences of including feedback as part of an assessment instrument. These aspects have been shown to have a interdependent effect that can be positive or negative (Nicol & Macfarlane-Dick, 2006).

In reviewing established feedback models, Boud and Molloy (2013) argued that they are at times based on unrealistic assumptions about the students and the educational setting. Commonly, due to resource constraints, the proposed feedback models or at least the mechanism for generating non-evaluative, supportive, timely, and specific feedback for each student is impractical or at least not sustainable in contemporary educational scenarios. At this juncture, LA/EDM work can play a significant role in moving feedback from an irregular and unidirectional state to an active dialogue between agents.

DATA-SUPPORTED FEEDBACK

The first initiatives using vast amounts of data to improve aspects of learning can be traced to areas such as *adaptive hypermedia* (Brusilovsky, 1996; Kobsa, 2007), *intelligent tutoring systems* (ITSs) (Corbett, Koedinger, & Anderson, 1997; Graesser, Conley, & Olney, 2012), and academic analytics (Baepler & Murdoch, 2010; Campbell, DeBlois, & Oblinger, 2007; Goldstein & Katz, 2005). Much of this research has taken place within LA/EDM research communities that share a common interest in data-intensive approaches to the research of educational setting, with the purpose of advancing educational practices (Baker & Inventado, 2014). While these communities have many similarities, there are some acknowledged differences between LA and EDM (Baker & Siemens, 2014). For example, EDM has a more reductionist focus on automated methods for discovery, as opposed to LA's human-led explorations situated within holistic systems. Baker and Inventado (2014) noted that the main differences between LA and EDM are not so much in the preferred methodologies, but in the focus, research questions, and eventual use of models.

When considering LA/EDM through the lens of feedback, the research approaches differ in relation to the direction and recipient of feedback. For instance, LA initiatives generally provide feedback aimed towards developing the student in the learning process (e.g., self-regulation, goal setting, motivation, strategies, and tactics). In contrast, EDM initiatives tend to focus on the provision of feedback to address changes in the learning environment (e.g., providing hints that modify a task, recommending heuristics that populate the environment with the relevant resources, et cetera). It is important to note that these generalizations are not a hard categorization between the communities, more so an observed trend in LA/EDM works that reflects their disciplinary backgrounds and interests. The following section further unpacks the work in both the EDM and LA communities related to the provision of feedback to aid student learning.

Approaches to Feedback in Educational Data Mining

Research undertaken in EDM is well connected and related to disciplines such as *artificial intelligence in education* (AIED) and *intelligent tutoring systems* (ITSs) (Pinkwart, 2016). Regarding feedback processes, a considerable number of EDM research initiatives have been concerned with developing and evaluating the effect of adapted and personalized feedback or recommendations to learners (Hegazi & Abugroon, 2016). This work is grounded on student modelling and/or predictive modelling research. Essentially, the focus has been on creating specific systems that can adapt the provision of feedback in order to respond to a student's particular needs, thereby facilitating improvements in learning, reinforcing (favourable) academic performance, or restraining students from performing certain behaviours (Romero & Ventura, 2013).

EDM approaches dealing with the provision of feedback have generally emphasized task-level feedback, with some notable exceptions (e.g., Arroyo, Meheranian, & Woolf, 2010; Kinnebrew & Biswas, 2012; Lewkow, Zimmerman, Riedesel, & Essa, 2015; Madhyastha & Tanimoto, 2009). Early research on EDM (see the EDM conference proceedings of 2008 and 2009) showcased a wide range of approaches aimed at providing feedback to learners through data-driven modelling (e.g., Mavrikis, 2008), learning-by-teaching agents (e.g., Jeong & Biswas, 2008), the provision of on-demand and instant prompts (Lynch, Ashley, Aleven, & Pinkwart, 2008), elaborated feedback as part of assessment tasks (Pechenizkiy, Calders, Vasilyeva, & De Bra, 2008), delayed feedback (Feng, Beck, & Heffernan, 2009), and process modelling (Pechenizkiy, Trcka, Vasilyeva, van der Aalst, & De Bra, 2009). This strand of EDM work includes the forward-oriented efforts for building an understanding of how such models can be enhanced to instrumentalize feedback mechanisms for informing future systems. In other words, algorithms could potentially provide the know-how to influence the design of new systems that provide better feedback. For instance, Barker-Plummer, Cox, and Dale (2009) suggested a need to move beyond the provision of better algorithms and understand how task-level feedback is influenced by the epistemic and pedagogical situation. In other words, the feedback at the level of learning process, or information about self-regulation skills, can help frame feedback at the task level.

A large portion of the studies related to adaptive feedback have been developed through intelligent tutoring systems (ITSs; e.g., Abbas & Sawamura, 2009; Eagle & Barnes, 2013; Feng et al., 2009), learning management systems (LMS; e.g., Dominguez, Yacef, & Curran, 2010; Lynch et al., 2008; Pechenizkiy et al., 2008), or equivalent single-user systems that provide a set of learning tasks to students in specific knowledge domains. Most of these systems capture student models in different ways: from traces of student behaviour, knowledge, achievement, cognitive states, or affective states for example. Based on these models, the system commonly offers various types of task-level feedback, such as next-step hints (e.g., Peddycord, Hicks, & Barnes, 2014); correctness hints, also known as flag feedback (Barker-Plummer, Cox, & Dale, 2011); positive or encouraging hints (Stefanescu, Rus, & Graesser, 2014); recommendations on next steps or tasks (Ben-Naim, Bain, & Marcus, 2009); or various combinations of the above. Hence, studies into behaviour modelling have been integral for developing automated feedback processes in EDM research (DeFalco, Baker, & D'Mello, 2014).

In recent years, EDM work in student modelling has been enriched by the emergence of new methods allowing researchers to generate feedback mechanisms for less structured learning tasks. An example includes the provision of formative and summative feedback on student writing (Allen & McNamara, 2015; Crossley, Kyle, McNamara, & Allen, 2014). The emergence of more sophisticated sensing devices and predictive algorithms has allowed the enhancement of student models by including traces of more complex human dimensions such as confidence, attitude, personality, motivation (Ezen-Can & Boyer, 2015), and affect (Fancsali, 2014). These more nuanced data aid the development of better responsive adaptive feedback mechanisms that can be personalized for each student. In parallel with the sophistication of student models, some researchers explored the notion of open learner modelling (OLM; Bull & Kay, 2016). The notion of OLM is similar to that of visual data representations but applied to the model built by a tool. OLMs originated within the AIED community in pursuit of providing less prescriptive forms of feedback compared with recommendations, corrective actions, or next-step hints. OLMs have gained renovated interest, as they allow the user (learner, teacher, peers, et cetera) to view and reflect on (or even scrutinize) the content of the learner model presented in human understandable forms. One of the advantages of these models is to help learners reflect and encourage self-regulating skills.

Recently, scaling up feedback gained traction in scholarly EDM work due to the increasing popularity of massive open online courses (MOOCs; Wen, Yang, & Rosé, 2014). Besides providing personalized feedback for student work in MOOCs (Pardos, Bergner, Seaton, & Pritchard, 2013), there is an interest in generating mechanisms to enable fair access to high-quality feedback in large cohorts. Some feedback solutions are addressing complex, open-ended learning tasks, building upon peer feedback (Piech et al., 2013) or through the provision of video-based feedback (Ostrow & Heffernan, 2014).

Although there has been a major emphasis in EDM to provide task-level, real-time feedback to students, other approaches have also been explored. For example, some efforts have focused on providing delayed feedback to avoid interruptions in students' learning processes (Feng et al., 2009; Johnson & Zaïane, 2012). There has also been interest in EDM to go beyond "corrective" feedback and understand the role that the polarity (positive, negative, or combined feedback) and the timing of feedback can play in students' dialogue (Ezen-Can & Boyer, 2013), in confidence (Lang, Heffernan, Ostrow, & Wang, 2015), or in collaborative scenarios (Olsen, Aleven, & Rummel, 2015). Providing feedback systematically targeting different levels of student activity is yet to receive due attention, though some examples have been offered. For instance, in Arroyo et al. (2010) digital learning companions acted as peers that provided feedback at cognitive (hints), affective (e.g., praise), and metacognitive levels (e.g., showing progress). The cognitive level, or the provision of hints, was offered at the task level. Showing progress addressed the capacity of self-reflection (i.e., monitor progress towards a goal). Other examples of feedback addressing regulation of learning have focused on supporting SRL behaviour and self-assessment (Bouchet, Azevedo, Kinnebrew, & Biswas, 2012); scaffolding high-level students strategies (Eagle & Barnes, 2014); recommending strategies of knowledge construction (Kinnebrew & Biswas, 2012); and understanding how feedback sits in students' learning processes (Howard, Johnson, & Neitzel, 2010).

Approaches to Feedback in Learning Analytics

Within the research in LA, a focus on feedback is generally interpreted as the need to communicate a student's state of learning to various stakeholders, i.e., teachers, students, or administrators. Early LA research (e.g., LAK 2011 and 2012 conference proceedings) did not focus on feedback per se, but emphasized LA as a discipline that needed to close the loop via scalable feedback processes (Clow, 2012; Lonn, Aguilar, & Teasley, 2013) to produce "actionable intelligence" (McKay, Miller, & Tritz, 2012). LA research recognized that feedback is conveyed through a multitude of disciplinary voices to humans with varying understandings of the agency and nature of learning (Suthers & Verbert, 2013). In line with that, Wise (2014) urged the design of data-driven learning interventions with awareness of how they are situated in their respective sociocultural contexts, and with the specific aim of addressing student support. Due to the significance of the context, perception and interpretation of data-supported feedback has been a distinct theme within LA feedback-related research. The LA community has searched for evidence and practices to ensure that the dialogue between the analytics and the stakeholders is taking place as imagined by the researchers. For instance, Corrin and de Barba (2015) inquired into student perceptions of dashboards; Beheshitha, Hatala, Gašević, and Joksimović (2016) examined if students with different achievement goal orientations perceived dashboard feedback in the same way; and a few studies investigated ways of making generated research more meaningful by combining qualitative interviews or the work of human interpreters with the data-driven analyses (Arnold, Lonn, & Pistilli, 2014; Clow, 2014; Mendiburo, Sulcer, & Hasselbring, 2014; Pardo, Ellis, & Calvo, 2015). The exposure of learners to some form of summary or indicators of their activity cannot be connected with a concrete level of feedback in the taxonomy proposed by Hattie and Timperley (2007). However, dashboards usually contain task level information, as inferring information about the learning process or self-regulation skills is much more challenging.

Similar to EDM, the interest of the LA community is in the provision of automated, scaled and real-time feedback to learners for self-monitoring and self-regulation processes, the third level in the taxonomy proposed by Hattie and Timperley (2007). Such direction has been well-captured through a steady growth of LA applications as tools for visualization, reflection, and awareness (e.g.,Verbert, Duval, Klerkx, Govaerts, & Santos, 2013; Verbert et al., 2014). Although specific task-level feedback is of less prominence than in EDM/ITS approaches, LA emphasizes more of the human-agency involved in interpreting and acting upon feedback. LA tends to promote process-level feedback by visualizing traces of learning activities. For instance, learning dashboards capture data sources, such as time spent, resources used, or social interaction, to enable learners to define goals and track progress towards these goals (for further review see Verbert et al., 2014). Recent applications of learning dashboards are shifting from the count of time or use of learning-related objects to visualizing progress related to a conceptualized process, e.g., table-top visualizations for inquiry-based learning (Charleer, Klerkx, & Duval, 2015), or visualizations of learning paths within competence graphs (Kickmeier-Rust, Steiner, & Dietrich, 2015). Visualizations informed by social network analysis (Dawson, 2010; Dawson, Bakharia, & Heathcote, 2010), as a part of social learning analytics (e.g., Ferguson & Buckingham Shum, 2012), remain a popular type of feedback on the social interaction process. These have been recently extended to help learners reflect on who they talk to, or where they are positioned in learner networks "in the wild," i.e., in distributed social media, such as Twitter or Facebook, and beyond the LMS (e.g., Bakharia, Kitto, Pardo, Gašević, & Dawson, 2016). Such network visualizations have also been offered to groups as representations of collective knowledge construction.

Feedback aimed towards developing student self-regulated learning proficiency is in its infancy. A promising approach to formative feedback embraces the *self* and various aspects of the learning process to support the development of resilient learner agency (Deakin Crick, Huang, Ahmed Shafi, & Goldspink, 2015). Another recent development includes the provision of feedback to students about their affective states. Grawemeyer et al. (2016) noted that students receiving affect-aware feedback were less bored and more consistently on-task than a comparative peer group receiving feedback only related to their performance. In essence, the authors demonstrate that the automated provision of feedback relating to a student's affective state can aid engagement and on-task behaviour. Ruiz et al. (2016) developed a visual dashboard providing visual feedback about student emotions and their evolution throughout the course. In this instance, the authors used the provision of self-reported emotional states as a source of self-reflection to improve performance and course designs. However, these studies also demonstrate that any noted success appears to be largely dependent on the learners' competence to self-regulate using the feedback from such learning analytics applications. Less reliance on the assumed level of students' competence is found when learning design or technology affordances prompt learner reflection. That is, learner thinking is externalized through writing text or annotations (also in-video), and formative feedback to this written text may then be offered.

The provision of feedback on written text, beyond essay grading, has been tackled by various initiatives in the area of discourse-centred analytics (De Liddo, Buckingham Shum, & Quinto, 2011). Also referred to as writing analytics, this area has a strong presence across the LA/EDM communities, with a significant overlap between methods for automatic text analysis, discourse analysis, and computational linguistics used to identify written text indicative of learning or knowledge construction (e.g., Simsek, Shum, De Liddo, Ferguson, & Sándor, 2014). In short, discourse-centred analytics offers feedback regarding the quality of cognitive engagement, or specifically assisting with aspects of writing as a domain skill, e.g., the quality of insight, genre, and so on (e.g., Crossley, Allen, Snow, & McNamara, 2015; Snow, Allen, Jacovina, Perret, & McNamara, 2015; Whitelock, Twiner, Richardson, Field, & Pulman, 2015). A noteworthy emergent trend within LA research emphasizes analysis of reflective writing (Buckingham Shum et al., 2016; Gibson & Kitto, 2015) offering formative feedback on learner's competency to reflect, potentially deepening individual engagement with both the content and process of learning.

CONCLUSION

This chapter has positioned one of the most influential aspects in the quality of the student learning experience, feedback, within the current research space of the EDM and LA communities. Despite the direct link between feedback and personalized learning, there are still significant gaps to be addressed. A dearth of research explores how students interact with and are transformed by algorithm-produced feedback. Furthermore, the relationship between the type of interventions that can be derived from data analysis and adequate forms of feedback remains inadequately explored. There is substantial literature analyzing the effect of feedback in learning experiences, but the area needs to be revisited with comprehensive data sets derived from technology mediation in learning experiences. In conventional face-to-face and blended learning scenarios, the increase in workload and limited instructor time are affecting the quality of feedback received by students. New emerging scenarios such as MOOCs pose significant challenges in providing high quality feedback to large student cohorts. LA and EDM are exploring how to address these limitations and propose new paradigms in which feedback is both scalable and effective. Although the initiatives in both communities have a strong connection with feedback, they differ in the areas of focal interest within which each discipline is devising its solutions. These foci are complementary, and often build upon each other. Consequently, both disciplines can benefit from a more comprehensive view of the role that feedback plays in a generic learning scenario, the elements involved, and the ultimate goal of prompting changes in the students' knowledge, beliefs, and attitudes. Practitioners from both research communities could well benefit from adopting a more comprehensive framework for feedback that supports a more effective integration across disciplines as well as the combination of humans and technology.

REFERENCES

Abbas, S., & Sawamura, H. (2009). Developing an argument learning environment using agent-based ITS (ALES). In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the 2nd International Conference on Educational Data Mining* (EDM2009), 1–3 July 2009, Cordoba, Spain (pp. 200–209). International Educational Data Mining Society.

Allen, L. K., & McNamara, D. S. (2015). You are your words: Modeling students' vocabulary knowledge with natural language processing tools. In O. C. Santos et al. (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 258–265). International Educational Data Mining Society.

Arnold, K. E., Lonn, S., & Pistilli, M. D. (2014). An exercise in institutional reflection: The learning analytics readiness instrument (LARI). *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 163–165). New York: ACM. doi:10.1145/2567574.2567621

Arroyo, I., Meheranian, H., & Woolf, B. P. (2010). Effort-based tutoring: An empirical approach to intelligent tutoring. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (EDM2010), 11–13 June 2010, Pittsburgh, PA, USA (pp. 1–10). International Educational Data Mining Society.

Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning*, 4(2). doi:10.20429/ijsotl.2010.040217

Baker, R., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), Learning analytics: From research to practice (pp. 61–75). Springer.

Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*. Cambridge University Press.

Bakharia, A., Kitto, K., Pardo, A., Gašević, D., & Dawson, S. (2016). Recipe for success: Lessons learnt from using xAPI within the connected learning analytics toolkit. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 378–382). New York: ACM.

Barker-Plummer, D., Cox, R., & Dale, R. (2009). Dimensions of difficulty in translating natural language into first order logic. In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the 2nd International Conference on Educational Data Mining* (EDM2009), 1–3 July 2009, Cordoba, Spain (pp. 220–229). International Educational Data Mining Society.

Barker-Plummer, D., Cox, R., & Dale, R. (2011). Student translations of natural language into logic: The grade grinder corpus release 1.0. In M. Pechenizkiy, T. Calders, C. Conati, S. Ventura, C. Romero, & J. Stamper (Eds.), *Proceedings of the 4th Annual Conference on Educational Data Mining* (EDM2011), 6–8 July 2011, Eindhoven, The Netherlands (pp. 51–60). International Educational Data Mining Society.

Beheshitha, S. S., Hatala, M., Gašević, D., & Joksimović, S. (2016). The role of achievement goal-orientations when studying effect of learning analytics visualisations. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 54–63). New York: ACM.

Ben-Naim, D., Bain, M., & Marcus, N. (2009). A user-driven and data-driven approach for supporting teachers in reflection and adaptation of adaptive tutorials. In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the 2nd International Conference on Educational Data Mining* (EDM2009), 1–3 July 2009, Cordoba, Spain (pp. 21–30). International Educational Data Mining Society.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. doi:10.1080/0969595980050102

Bouchet, F., Azevedo, R., Kinnebrew, J. S., & Biswas, G. (2012). Identifying students' characteristic learning behaviors in an intelligent tutoring system fostering self-regulated learning. In K. Yacef, O. Zaïane, A. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (EDM2012), 19–21 June, 2012, Chania, Greece (pp. 65–72). International Educational Data Mining Society.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*(2), 151–167. doi:10.1080/713695728

Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education,* 38(6), 698–712. doi:10.1080/02602938.2012.691462

Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User-Adapted Interaction,* 6, 87–129.

Buckingham Shum, S., Sandor, A., Goldsmith, R., Wang, X., Bass, R., & McWilliams, M. (2016). Reflecting on reflective writing analytics: Assessment challenges and iterative evaluation of a prototype tool. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 213–222). New York: ACM.

Bull, S., & Kay, J. (2016). SMILI : A framework for interfaces to learning data in open learner models, learning analytics and related fields. *International Journal of Artificial Intelligence in Education, 26*(1), 293–331.

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research,* 65(3), 245–281.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review*, 42, 40–57.

Charleer, S., Klerkx, J., & Duval, E. (2015). Exploring inquiry-based learning analytics through interactive surfaces. *Proceedings of the Workshop on Visual Aspects of Learning Analytics* (VISLA '15), 16–20 March 2015, Poughkeepsie, NY, USA. http://ceur-ws.org/Vol-1518/paper6.pdf

Clow, D. (2012). The learning analytics cycle: Closing the loop effectively. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 134–138). New York: ACM. doi:10.1145/2330601.2330636

Clow, D. (2014). Data wranglers: Human interpreters to help close the feedback loop. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 49–53). New York: ACM. doi:10.1145/2567574.2567603

Corbett, A. T., Koedinger, K. R., & Anderson, J. R. (1997). Intelligent tutoring systems. In M. Heander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human–computer interaction* (2nd ed., pp. 849–870): Elsevier Science B. V.

Corrin, L., & de Barba, P. (2015). How do students interpret feedback delivered via dashboards? *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 430–431). New York: ACM. doi:10.1145/2723576.2723662

Crossley, S., Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Pssst... textual features... there is more to automatic essay scoring than just you! *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 203–207). New York: ACM. doi:10.1145/2723576.2723595

Crossley, S., Kyle, K., McNamara, D. S., & Allen, L. (2014). The importance of grammar and mechanics in writing assessment and instruction: Evidence from data mining. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 300–303). International Educational Data Mining Society.

Dawson, S. (2010). "Seeing" the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, *41*(5), 736–752.

Dawson, S., Bakharia, A., & Heathcote, E. (2010). SNAPP: Realising the affordances of real-time SNA within networked learning environments. In L. Dirckinck-Holmfeld, V. Hodgson, C. Jones, M. de Laat, D. McConnell, & T. Ryberg (Eds.), *Proceedings of the 7th International Conference on Networked Learning* (NLC 2010), 3–4 May 2010, Aalborg, Denmark. http://www.networkedlearningconference.org.uk/past/nlc2010/abstracts/PDFs/Dawson.pdf

De Liddo, A., Buckingham Shum, S., & Quinto, I. (2011). Discourse-centric learning analytics. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 23–33). New York: ACM.

Deakin Crick, R., Huang, S., Ahmed Shafi, A., & Goldspink, C. (2015). Developing resilient agency in learning: The internal structure of learning power. *British Journal of Educational Studies*, 63(2), 121–160. doi:10.1080/00071005.2015.1006574

DeFalco, J. A., Baker, R., & D'Mello, S. K. (2014). Addressing behavioral disengagement in online learning. *Design recommendations for intelligent tutoring systems* (Vol. 2, pp. 49–56). Orlando, FL: U.S. Army Research Laboratory.

Dominguez, A. K., Yacef, K., & Curran, J. R. (2010). Data mining for generating hints in a python tutor. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (EDM2010), 11–13 June 2010, Pittsburgh, PA, USA (pp. 91–100). International Educational Data Mining Society.

Eagle, M., & Barnes, T. (2013). Evaluation of automatically generated hint feedback. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July 2013, Memphis, TN, USA (pp. 372–374). International Educational Data Mining Society/Springer.

Eagle, M., & Barnes, T. (2014). Data-driven feedback beyond next-step hints. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 444–446). International Educational Data Mining Society.

Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of Educational Research*, 83(1), 70–120. doi:10.3102/0034654312474350

Ezen-Can, A., & Boyer, K. E. (2013). Unsupervised classification of student dialogue acts with query-likelihood clustering. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July 2013, Memphis, TN, USA (pp. 20–27). International Educational Data Mining Society/Springer.

Ezen-Can, A., & Boyer, K. E. (2015). Choosing to interact: Exploring the relationship between learner personality, attitudes, and tutorial dialogue participation. In O. C. Santos et al. (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 125–128). International Educational Data Mining Society.

Fancsali, S. E. (2014). Causal discovery with models: Behavior, affect, and learning in cognitive tutor algebra. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 28–35). International Educational Data Mining Society.

Feng, M., Beck, J. E., & Heffernan, N. T. (2009). Using learning decomposition and bootstrapping with randomization to compare the impact of different educational interventions on learning. In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the 2nd International Conference on Educational Data Mining* (EDM2009), 1–3 July 2009, Cordoba, Spain (pp. 51–60). International Educational Data Mining Society.

Ferguson, R., & Buckingham Shum, S. (2012). Social learning analytics: Five approaches. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 23–33). New York: ACM.

Gibson, A., & Kitto, K. (2015). Analysing reflective text for learning analytics: An approach using anomaly recontextualisation. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 51–60). New York: ACM. doi:10.1145/2723576.2723635

Goldstein, P. J., & Katz, R. N. (2005). *Academic analytics: The uses of management information and technology in higher education.* Educause Center for Applied Research.

Graesser, A. C., Conley, M. W., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook* (pp. 451–473). American Psychological Association.

Grawemeyer, B., Mavrikis, M., Holmes, W., Guitierrez-Santos, S., Wiedmann, M., & Rummel, N. (2016). Affective off-task behaviour: How affect-aware feedback can improve student learning. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 104–113). New York: ACM.

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses related to achievement*. New York: Routledge.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. doi:10.3102/003465430298487

Hegazi, M. O., & Abugroon, M. A. (2016). The state of the art on educational data mining in higher education. *International Journal of Computer Trends and Technology*, 31, 46–56.

Howard, L., Johnson, J., & Neitzel, C. (2010). Examining learner control in a structured inquiry cycle using process mining. In R. S. J. d. Baker, A. Merceron, & P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (EDM2010), 11–13 June 2010, Pittsburgh, PA, USA (pp. 71–80). International Educational Data Mining Society.

Jeong, H., & Biswas, G. (2008). Mining student behavior models in learning-by-teaching environments. In R. S. J. d. Baker, T. Barnes, & J. E. Beck (Eds.), *Proceedings of the 1st International Conference on Educational Data Mining* (EDM'08), 20–21 June 2008, Montreal, QC, Canada (pp. 127–136). International Educational Data Mining Society.

Johnson, S., & Zaïane, O. R. (2012). Deciding on feedback polarity and timing. In K. Yacef, O. Zaïane, A. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (EDM2012), 19–21 June, 2012, Chania, Greece (pp. 220–221). International Educational Data Mining

Society.

Kickmeier-Rust, M., Steiner, C. M., & Dietrich, A. (2015). Uncovering learning processes using competence-based knowledge structuring and Hasse diagrams. *Proceedings of the Workshop on Visual Aspects of Learning Analytics* (VISLA '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 36–40).

Kinnebrew, J. S., & Biswas, G. (2012). Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. In K. Yacef, O. Zaïane, A. Hershkovitz, M. Yudelson, & J. Stamper (Eds.), *Proceedings of the 5th International Conference on Educational Data Mining* (EDM2012), 19–21 June, 2012, Chania, Greece (pp. 57–64). International Educational Data Mining Society.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284.

Kobsa, A. (2007). Privacy-enhanced web personalization. In P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), *The adaptive web: Methods and strategies of web personalization* (pp. 628–670). Springer. doi:10.1007/978-3-540-72079-9_4.

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, *1*(4), 279–308.

Lang, C., Heffernan, N., Ostrow, K., & Wang, Y. (2015). The impact of incorporating student confidence items into an intelligent tutor: A randomized controlled trial. In O. C. Santos et al. (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 144–149). International Educational Data Mining Society.

Lewkow, N., Zimmerman, N., Riedesel, M., & Essa, A. (2015). Learning analytics platform: Towards an open scalable streaming solution for education. In O. C. Santos et al. (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 460–463). International Educational Data Mining Society.

Lockyer, L., Heathcote, E., & Dawson, S. (2013). Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist*, *57*(10), 1439–1459. doi:10.1177/0002764213479367

Lonn, S., Aguilar, S., & Teasley, S. D. (2013). Issues, challenges, and lessons learned when scaling up a learning analytics intervention. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 235–239). New York: ACM. doi:10.1145/2460296.2460343

Lynch, C., Ashley, K. D., Aleven, V., & Pinkwart, N. (2008). Argument graph classification via genetic programming and C4.5. In R. S. J. d. Baker, T. Barnes, & J. E. Beck (Eds.), *Proceedings of the 1st International Conference on Educational Data Mining* (EDM'08), 20–21 June 2008, Montreal, QC, Canada (pp. 137–146). International Educational Data Mining Society.

Madhyastha, T. M., & Tanimoto, S. (2009). Student consistency and implications for feedback in online assessment systems. In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the 2nd International Conference on Educational Data Mining* (EDM2009), 1–3 July 2009, Cordoba, Spain (pp. 81–90). International Educational Data Mining Society.

Mavrikis, M. (2008). Data-driven modelling of students' interactions in an ILE. In R. S. J. d. Baker, T. Barnes, & J. E. Beck (Eds.), *Proceedings of the 1st International Conference on Educational Data Mining* (EDM'08), 20–21 June 2008, Montreal, QC, Canada (pp. 87–96). International Educational Data Mining Society.

McKay, T., Miller, K., & Tritz, J. (2012). What to do with actionable intelligence: E2Coach as an intervention engine. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 88–91). New York: ACM. doi:10.1145/2330601.2330627

Mendiburo, M., Sulcer, B., & Hasselbring, T. (2014). Interaction design for improved analytics. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 78–82). New York: ACM. doi:10.1145/2567574.2567628

Mory, E. H. (2004). Feedback research revisited. In D. H. Jonassen (Ed.), *Handbook of research on educational communication and technology* (Vol. 2, pp. 745–783). Mahwah, NJ: Lawrence Erlbaum.

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218. doi:10.1080/03075070600572090

Olsen, J. K., Aleven, V., & Rummel, N. (2015). Predicting student performance in a collaborative learning environment. In O. C. Santos et al. (Eds.), *Proceedings of the 8th International Conference on Educational Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 211–217). International Educational Data Mining Society.

Ostrow, K. S., & Heffernan, N. T. (2014). Testing the multimedia principle in the real world: A comparison of video vs. text feedback in authentic middle school math assignments. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 296–299). International Educational Data Mining Society.

Pardo, A., Ellis, R. A., & Calvo, R. A. (2015). Combining observational and experiential data to inform the redesign of learning activities. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 305–309). New York: ACM. doi:10.1145/2723576.2723625

Pardos, Z. A., Bergner, Y., Seaton, D. T., & Pritchard, D. E. (2013). Adapting Bayesian knowledge tracing to a massive open online course in edX. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July 2013, Memphis, TN, USA (pp. 137–144). International Educational Data Mining Society/Springer.

Pechenizkiy, M., Calders, T., Vasilyeva, E., & De Bra, P. (2008). Mining the student assessment data: Lessons drawn from a small scale case study. In R. S. J. d. Baker, T. Barnes, & J. E. Beck (Eds.), *Proceedings of the 1st International Conference on Educational Data Mining* (EDM'08), 20–21 June 2008, Montreal, QC, Canada (pp. 187–191). International Educational Data Mining Society.

Pechenizkiy, M., Trcka, N., Vasilyeva, E., van der Aalst, W., & De Bra, P. (2009). Process mining online assessment data. In T. Barnes, M. Desmarais, C. Romero, & S. Ventura (Eds.), *Proceedings of the 2nd International Conference on Educational Data Mining* (EDM2009), 1–3 July 2009, Cordoba, Spain (pp. 279–288). International Educational Data Mining Society.

Peddycord III, B., Hicks, A., & Barnes, T. (2014). Generating hints for programming problems using intermediate output. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 92–98). International Educational Data Mining Society.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned models of peer assessment in MOOCs. In S. K. D'Mello, R. A. Calvo, & A. Olney (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July 2013, Memphis, TN, USA (pp. 153–160). International Educational Data Mining Society/Springer.

Pinkwart, N. (2016). Another 25 years of AIED? Challenges and opportunities for intelligent educational technologies of the future. *International Journal of Artificial Intelligence in Education*, 26(2), 771–783. doi:10.1007/s40593-016-0099-7

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self-regulated learning. *International Journal of Educational Research*, 31, 459–470. doi:10.1016/S0883-0355(99)00015-4

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27. doi:10.1002/widm.1075

Ruiz, S., Charleer, S., Urretavizcaya, M., Klerkx, J., Fernandez-Castro, I., & Duval, E. (2016). Supporting learning by considering emotions: Tracking and visualisation. A case study. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 254–263). New York: ACM. doi:10.1145/2883851.2883888

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. doi:10.3102/0034654307313795

Simsek, D., Shum, S. B., De Liddo, A., Ferguson, R., & Sándor, Á. (2014). Visual analytics of academic writing. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 265–266). New York: ACM. doi:10.1145/2567574.2567577

Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). You've got style: Detecting writing flexibility across time. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 194–202). New York: ACM. doi:10.1145/2723576.2723592

Stefanescu, D., Rus, V., & Graesser, A. C. (2014). Towards assessing students' prior knowledge from tutorial dialogues. In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 197–201). International Educational Data Mining Society.

Suthers, D., & Verbert, K. (2013). Learning analytics as a "middle space." *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 1–4). New York: ACM. doi:10.1145/2460296.2460298

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, 57(10), 1500–1509.

Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Assche, F., Parra, G., & Klerkx, J. (2014). Learning dashboards: An overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6), 1499–1514. doi:10.1007/s00779-013-0751-2

Wen, M., Yang, D., & Rosé, C. P. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 130–137). International Educational Data Mining Society.

Whitelock, D., Twiner, A., Richardson, J. T. E., Field, D., & Pulman, S. (2015). OpenEssayist: A supply and demand learning analytics tool for drafting academic essays. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 208–212). New York: ACM. doi:10.1145/2723576.2723599

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, 11(1), 49–65. doi:10.1080/0969594042000208994

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition and Learning*, 9(2), 229–237. doi:10.1007/s11409-014-9113-3

Winne, P. H., & Baker, R. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, 5(1), 1–8.

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 203–211). New York: ACM.

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17. doi:10.1207/s15326985ep2501_2

Chapter 15: Epistemic Network Analysis: A Worked Example of Theory-Based Learning Analytics

David Williamson Shaffer and A. R. Ruis

Wisconsin Center for Education Research, University of Wisconsin–Madison, USA
DOI: 10.18608/hla17.015

ABSTRACT

In this article, we provide a worked example of a theory-based approach to learning analytics in the context of an educational game. We do this not to provide an ideal solution for others to emulate, but rather to explore the affordances of a theory-based - rather than data-driven - approach. We do so by presenting 1) epistemic frame theory as an approach to the conceptualization of learning; 2) data from an epistemic game, an approach to educational game design based on epistemic frame theory; and 3) epistemic network analysis (ENA), a technique for analyzing discourse and other data for evidence of complex thinking based on the same theory. We describe ENA through a specific analytic result, but our aim is to explore how this result exemplifies what we consider a key "best practice" in the field of learning analytics.

Keywords: Epistemic frame theory, epistemic game, epistemic network analysis (ENA), evidence centred design

In this chapter, we look at the role of theory in learning analytics. Researchers who study learning are blessed with unprecedented quantities of data, whether information about staggeringly large numbers of individuals or data showing the microscopic, moment-by-moment actions in the learning process. It is a brave new world. We can look at second-by-second changes in where students focus their attention, or examine what study skills are effective by looking at thousands of students in a MOOC.

As Wise and Shaffer (2016) argue in a special section of the *Journal of Learning Analytics*, however, it is dangerous to think that with enough information, the data can speak for themselves — that we can conduct analyses of learning without theories of learning. In fact, the opposite is true. With larger amounts of data, theory plays an even more critical role in analysis. Put in simple terms, most extant statistical tools were developed for datasets of a particular size and type: large enough so that random effects are normally distributed, but small enough to be obtained using traditional data collection techniques. Applying these techniques to datasets that are orders of magnitude larger in length and number of variables without a strong theoretical foundation is perilous at best.

In what follows, we look at this question not by analyzing the problems of applying statistics without a theoretical framework. What Wise and Shaffer suggest — and what the articles and commentaries in the special section of the *Journal of Learning Analytics* show — is that conducting theory-based learning analytics is challenging. As a result, our approach in what follows is to examine the role of theory in learning analytics through the use of a *worked example*: the presentation of a problem along with a step-by-step description of its solution (Atkinson, Derry, Renkl, & Wortham, 2000).

In doing so, our aim is not to provide an ideal solution for others to emulate, nor to suggest that our particular use of theory in learning analytics is better than others. Rather, our goal is to reflect on the importance of a theory-based approach — as opposed to an atheoretical or data-driven approach — to the analysis of large educational datasets. We do so by presenting *epistemic network analysis* (ENA; Andrist, Collier, Gleicher, Mutlu, & Shaffer, 2015; Arastoopour, Shaffer, Swiecki, Ruis, & Chesler, 2016; Chesler et al., 2015; Nash & Shaffer, 2013; Rupp, Gustha, Mislevy, & Shaffer, 2010; Rupp, Sweet, & Choi, 2010; Shaffer et al., 2009; Shaffer, Collier, & Ruis, 2016; Svarovsky, 2011), a novel learning analytic technique. But importantly, we present ENA in the context of *epistemic frame theory* – the approach to learning on which ENA was based – and apply it to data from an *epistemic game*, an approach to educational game design based on epistemic frame theory. We thus describe ENA through a specific analytic result to examine how this result exemplifies the alignment of theory, data, and analysis as a "best practice" in the field.

DATA

The data we will use to explore this particular worked example come from an epistemic game (Shaffer, 2006, 2007), a simulation of authentic professional practice that helps students learn to think in the way that experts do. Specifically, the data come from the epistemic game *Land Science*, an online urban planning simulation in which students assume the role of interns at a fictitious firm competing for a redevelopment contract from the city of Lowell, Massachusetts. They work in small teams, communicating via chat and email, to develop a rezoning plan for the city that addresses the demands of different stakeholder groups. To do this, students review research briefs and other resources, conduct a survey of stakeholder preferences, and model the effects of land-use changes on pollution, revenue, housing, and other indicators using a GIS mapping tool. Because no rezoning plan can meet all stakeholder preferences, students must justify the decisions they make in their final proposals.

Land Science has been used with high school students and first-year college students more than 30 times. Our prior research (Bagley & Shaffer, 2009, 2015b; Nash, Bagley, & Shaffer, 2012; Nash & Shaffer, 2012; Shaffer, 2007) has shown that Land Science helps students learn content and practices in urban ecology, urban planning, and related fields, and it also helps them develop skills, interests, and motivation to improve performance in school.

As with many educational technologies, Land Science records all of the things that students do during the simulation, including their chats and emails, their notebooks and other work products, and every keystroke and mouse-click. This makes it possible to analyze not only students' final products but also the problem-solving processes they use.

In the worked example presented below, we examine the chat conversations from 311 students who used the same version of *Land Science*, including seven groups of college students (n = 155), eight groups of high school students (n = 110), and three groups of gifted and talented high school students (n = 46). In its entirety, this dataset contains 44,964 lines of chat.

THEORY

Our analysis of the chat data from *Land Science* is informed by *epistemic frame theory* (Shaffer, 2004, 2006, 2007, 2012). The theory of epistemic frames models the ways of thinking, acting, and being in the world of some *community of practice* (Lave & Wenger, 1991; Rohde & Shaffer, 2004). A community of practice, or a group of people with a common approach to framing, investigating, and solving problems, has a repertoire of knowledge and skills, a set of values that guides how skills and knowledge should be used, and a set of processes for making and justifying decisions. A community also has a common identity exhibited both through overt markers and through the enactment of skills, values, and decision-making processes characteristic of the community.

Becoming part of a community of practice, in other words, means acquiring a particular Discourse: a way of "talking, listening, writing, reading, acting, interacting, believing, valuing, and feeling (and using various objects, symbols, images, tools, and technologies)" (Gee, 1999, p. 719). A Discourse is the manifestation of a culture and, based on Goodwin's (1994) *professional vision*, an epistemic frame is the grammar of a Discourse: a formal description of the configuration of Discourse elements exhibited by members of a particular community of practice.

Importantly, however, it is not mere *possession* of relevant knowledge, skills, values, practices, and other attributes that characterizes the epistemic frame of a community, but *the particular set and configuration of them*. The concept of a *frame* comes from Goffman (1974) (see also Tannen, 1993). Activity is interpreted in terms of a frame: the rules and premises that shape perceptions and actions, or the set of norms and practices by which experiences are interpreted. An epistemic frame is thus revealed by the actions and interactions of an individual engaged in authentic tasks (or simulations of authentic tasks).

To identify analytically the connections among elements that make up an epistemic frame, we identify *co-occurrences* of them in student discourse – in this case, in the conversations they have in an online chat program. Researchers (Chesler et al., 2015; Dorogovtsev & Mendes, 2013; i Cancho & Solé, 2001; Landauer, McNamara, Dennis, & Kintsch, 2007; Lund & Burgess, 1996) have shown that co-occurrences of concepts in a given segment of discourse data are a good indicator of cognitive connections, particularly when the co-occurrences are frequent (Newman, 2004). These concepts can be identified a priori from a theoretical or empirical analysis, or from an ethnographic study of the community in action.

ENA operationalizes epistemic frame theory by identifying co-occurrences in segments of discourse data and modelling the weighted structure of co-occurrences. ENA represents these patterns of co-occurrence in a dynamic network model that quantifies changes in the strength and composition of an epistemic frame over time – a process we describe in the next section.

ENA

ENA models the weighted structure of connections in discourse data, or in any kind of stanza-based interaction data. In what follows, we describe both the general principles of the ENA method and the specific process by which the current version of ENA software – www.epistemicnetwork.org – implements the ENA algorithms.

Stanza-Based Interaction Data

Before we describe how ENA operationalizes epistemic frame theory, it is important to understand how data is configured for analysis using ENA. Consider the simplified data in Table 15.1, which shows excerpts from two conversations held by one group of students in *Land Science*. In the five columns to the right are the concepts, or codes, whose pattern of association we want to model. In this case, the codes represent various aspects of professional urban planning practice – that is, various elements of an urban planning epistemic frame.

Note that sometimes we can see relations among the codes in a single utterance, as in In Line 3, where Jorge references knowledge of both social issues and environmental issues. In other cases, relations occur across utterances: in Line 10, Depesh talks about the trade-off involved in increasing open space, which responds to and builds on Natalie's more general comment about trade-offs in Line 8. However, we do not necessarily want to look at the relations among codes across *all* turns of talk. For example, two separate conversations are represented in Table 15.1. Both involve the same group of students (Group 3), but the conversations took place on two different days while the students were working on two different activities.

To create a network model of these data, we need to group the lines into stanzas. The key idea behind a stanza is that (a) codes in lines *anywhere within the*

Table 15.1. Edited Excerpt of Discourse Data Coded in ENA Format

Line	Activity	Group	Username	Created	Utterance	E.social.issues	S.zoning.codes	K.social.issues	K.zoning.codes	K.environment
1	VSV Meeting	3	Natalie	02/11/14 10:03	Okay, so what do the stakeholders want?	0	0	0	0	0
2	VSV Meeting	3	Depesh	02/11/14 10:03	talking w/ stakeholders, we learned that there are many issues within the city but there are some bariers that prevent these issues from being easily solved	0	0	1	0	0
3	VSV Meeting	3	Jorge	02/11/14 10:04	Yeah, the stakeholders care a lot about the environmental impact in the area as well as the need for low income housing	0	0	1	0	1
4	VSV Meeting	3	Depesh	02/11/14 10:04	they cared about different issues but they all wanted to create a healthy and livable community	0	0	1	0	0
5	VSV Meeting	3	Natalie	02/11/14 10:05	I agree. They are also worried about the quality of the water.	0	0	0	0	1
6	VSV Meeting	3	Jessie	02/11/14 10:06	and they want more housing opportunities for low-income residents	0	0	1	0	0
7	iPlan Meeting	3	Jorge	02/13/14 10:21	Quick question, what does the indicator P mean?	0	0	0	0	1
8	iPlan Meeting	3	Natalie	02/13/14 10:21	I found that certain indicators changed when altering the zoning designations of specific sites. Each change in zoning category came with its benefits and drawbacks. There was usually a tradeoff involved.	0	1	0	1	0
9	iPlan Meeting	3	Jessie	02/13/14 10:21	@Jorge: P = phosporous	0	0	0	0	1
10	iPlan Meeting	3	Depesh	02/13/14 10:22	yeah, if you add open space you can help run-off and nesting but hurt the job totals	1	0	1	0	1
11	iPlan Meeting	3	Jorge	02/13/14 10:25	Yeah, everything affects something.	0	0	0	0	0

same stanza are *related* to one another in the model, and (b) codes in lines that are *not in the same stanza* are *not related* to one another in the model. In this case, stanzas indicate which co-occurrences of concepts represent meaningful cognitive connections among the epistemic frame elements of urban planning.

ENA Models

To construct a network model from stanza-based interaction data, ENA *collapses* the stanzas. Usually this is done as a *binary* accumulation: if any line of data in the stanza contains code A, then the stanza contains code A. For example, the data shown in Table 1 would be collapsed as shown in Table 15.2 if we choose "Activity" to define the stanzas.

Table 15.2. Stanzas by Activity for Group 3

Activity	Group	E.social.issues	S.zoning.codes	K.social.issues	K.zoning.codes	K.environment
VSV Meeting	3	0	0	0	0	0
VSV Meeting	3	0	0	1	0	0

ENA then creates an *adjacency matrix* for each stanza, which summarizes the co-occurrence of codes (see Table 15.3). The diagonal of the matrix contains all zeros because codes in this model, and in general in ENA, do not co-occur with themselves. Each adjacency matrix, in this case, represents the connections that Group 3 made among urban planning epistemic frame elements during a particular activity. For example, in the VSV Meeting activity, K.social.issues, and K.environment both occurred in Group 3's discourse. The adjacency matrix representing that activity in Table 15.3 (left) thus contains a 1 in the cells that represent the co-occurrence of those two codes.

The adjacency matrices representing each stanza are then summed into a cumulative adjacency matrix for each unit of analysis in the dataset. The simple example shown in Table 15.3 would thus be represented by the cumulative adjacency matrix shown in Table 15.4. At the end of this process of accumulation, each unit in the dataset (in this case, each group) is associated with a cumulative adjacency matrix that represents the weighted pattern of co-occurrence (cognitive connections) among the codes (epistemic frame elements) for that unit.

To understand the structure of connections across different units – the relationships among their networks of connections, or the differences among their cumulative adjacency matrices – ENA represents each cumulative adjacency matrix as a vector in a high-dimensional space, where each vector is defined by the values in the upper diagonal half of the matrix. Note that the *dimensions* of this space correspond to the strength of association between every pair of codes.

Table 15.3. Stanzas by Activity for Group 3

Group 3 VSV Meeting	E.social.issues	S.zoning.codes	K.social.issues	K.zoning.codes	K.environment
E.social.issues	0	0	0	0	0
S.zoning.codes	0	0	0	0	0
K.social.issues	0	0	0	0	1
K.zoning.codes	0	0	0	0	0
K.environment	0	0	1	0	0

Group 3 iPlan Meeting	E.social.issues	S.zoning.codes	K.social.issues	K.zoning.codes	K.environment
E.social.issues	0	1	1	1	1
S.zoning.codes	1	0	1	1	1
K.social.issues	1	1	0	1	1
K.zoning.codes	1	1	1	0	1
K.environment	1	1	1	1	0

Table 15.4. Cumulative Adjacency Matrix for Group 3, Summing the Two Adjacency Matrices Shown in Table 3

Group 3	E.social.issues	S.zoning.codes	K.social.issues	K.zoning.codes	K.environment
E.social.issues	0	1	1	1	1
S.zoning.codes	1	0	1	1	1
K.social.issues	1	1	0	1	2
K.zoning.codes	1	1	1	0	1
K.environment	1	1	2	1	0

Before analyzing the data in ENA space, ENA divides each vector by its length to normalize the data. This is done because the *length* of a vector is potentially affected by the number of stanzas contained in the unit of analysis. More stanzas are likely to produce more co-occurrences, which result in longer vectors. This is problematic because two vectors may represent the same *pattern* of association, and thus point in the same *direction*, but represent different *numbers* of stanzas, and thus have different *lengths*.

Once the data are normalized, ENA performs a *singular value decomposition* (SVD), a projection that centres the data but does not rescale it. This maximizes the variance accounted for in the data (similar to a principal components analysis). However, unlike a traditional PCA or factor analysis, (a) ENA is performed on the co-occurrences from the cumulative adjacency matrices, rather than on the counts or strengths of the codes themselves, and (b) ENA performs a sphere or cosine norm on the original data and centres it, but does not rescale the dimensions individually.

Interpretation of ENA Models

Once an ENA model is created, a suite of tools can be used to understand and create a meaningful interpretation. For example, in the *Land Science* dataset described above, the chat utterances of all students were coded for 24 urban planning epistemic frame elements (see Appendix I) using a previously developed and validated automated coding process (Bagley & Shaffer, 2015b; Nash & Shaffer, 2011). Codes relevant to authentic urban planning practice were developed based on an ethnographic study of how urban planners are trained (Bagley & Shaffer, 2015a).

ENA models are typically visualized using two-dimensions at a time, which facilitates interpretation. Figure 15.1, for example, shows the cumulative epistemic network of a high school student (Student A) who participated in *Land Science*. The network models the structure of connections among the elements of the student's urban planning epistemic frame. In this case, Student A's network shows a number of connections among knowledge elements, such as knowledge of social issues and knowledge of complex systems; epistemological elements, such as compromise; and the skill of using urban planning tools (such as a preference survey). The network is also weighted: thicker, more saturated lines represent stronger connections, whereas thinner, less saturated lines represent weaker connections. The thickness/saturation of a line is proportional to the number of stanzas in which the connection between the two epistemic frame elements occurred.

While we can draw some conclusions about this student's network – for example, Student A made cognitive connections mostly among basic knowledge and skills – in many cases, the salient features of a network are easier to identify in comparison with other networks. Figure 15.2 shows the urban planning epistemic network of a second high school student (Student B). Like Student A, Student B made a number of connections among basic knowledge elements, but Student B's network exhibits more and stronger connections overall as well as connections to additional elements, most notably to more advanced skills, such as scientific thinking, and to epistemological attributes.

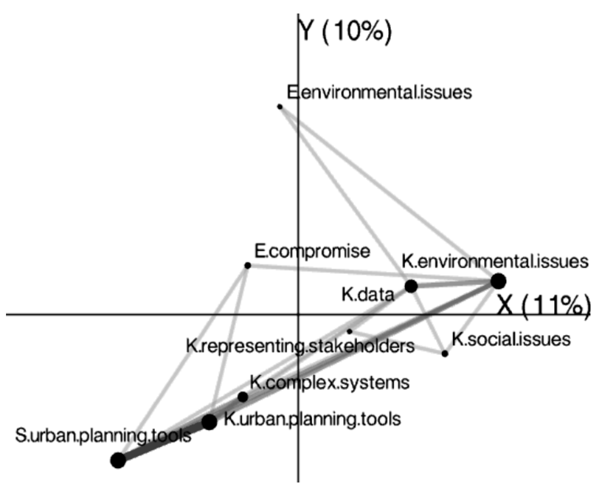


Figure 15.1. Epistemic network of a high school student (Student A) representing the structure of cognitive connections the student made while solving a simulated urban redevelopment problem. Percentages in parentheses indicate the total variance in the model accounted for by each dimension.the integration of multiple sources of data.

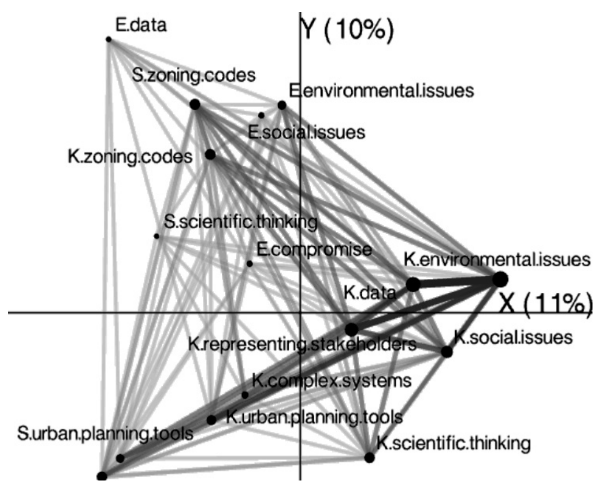


Figure 15.2. Epistemic network of a high school student (Student B) representing the cognitive connections the student made while solving a simulated urban redevelopment problem.

As discussed above, epistemic frame theory suggests that the epistemic frame of urban planning (or any community of practice) is defined by how and to what extent urban planning knowledge, skills, values, and other attributes are interconnected. In this example, ENA reveals that Student B's network is more overtly *epistemic*: she explained and justified her thinking in the way that urban planners do, and is thus learning to think like an urban planner.

What makes this comparison between Students A and B possible is that the nodes in both epistemic networks appear in exactly the same places in the network projection space – for these two students, and for all the students in the dataset. This invariance in node placement allows us to compare the network projections of different units directly, but this method of direct comparison only works for very small numbers of networks – what if we want to compare dozens or even hundreds of networks? For example, what if we want to compare all 110 high school students in this dataset, or compare the high school students with the college students? ENA makes this possible by representing each network as a single point in the projection space, such that each point is the *centroid* of the corresponding network.

The centroid of a network is similar to the centre of mass of an object. Specifically, the centroid of a network graph is the arithmetic mean of the edge weights of the network model distributed according to the network projection in space. The important point here is that *the centroid of an ENA network summarizes the network as a single point in the projection space that accounts for the weighted structure of connections in the specific arrangement of the network model*.

The locations of the nodes in the network projection are determined by an optimization routine to minimize, for any given network, the distance between (a) the centroid of the network graph, and (b) the point that represents the network under the SVD rotation. Choosing fixed node positions to have the centroid of a network correspond to the position of the network in a projected space allows for characterization of the projection space – and thus of the salient differences among different networks in the ENA model. In this case, we can interpret the projection space in the following way: toward the lower left are basic professional skills, such as professional communication and use of urban planning tools; toward the right are knowledge elements related to the specific redevelopment problem and to knowledge of more general topics, such as data and scientific thinking; and toward the upper left are elements of more advanced urban planning thinking, especially epistemological elements – making and justifying decisions according to urban planning conventions – and the use of zoning codes.

We can thus compare a large number of different networks simultaneously because centroids located in the same part of the projection space represent networks with similar patterns of connections, while centroids located in different parts of the projection space represent networks with different patterns of connections[1]. This allows us to explore any number of research questions about students' urban planning epistemic frames. One question we might ask of the *Land Science* dataset is *How do the epistemic networks of the different student populations (college, high school, and gifted high school) differ*? For example, when we plot the centroids of the college students and the high school students (Figure 15.3), the two groups are distributed differently. To determine if the difference is statistically significant, we can perform an independent samples t test on the mean positions of the two populations in the projection space. The college students (dark) and high school students (light) are significantly different on both dimensions:

$\bar{x}_{College} = -0.083$, $\bar{x}_{HS} = 0.115$, $t = -7.025$, $p < 0.001$, $Cohen's\ d = -0.428$

$\bar{y}_{College} = 0.040$, $\bar{y}_{HS} = -0.045$, $t = 3.199$, $p = 0.002$, $Cohen's\ d = 0.186$

When the gifted and talented high school students are included in the analysis, in some respects they are more similar to the college students, and in others

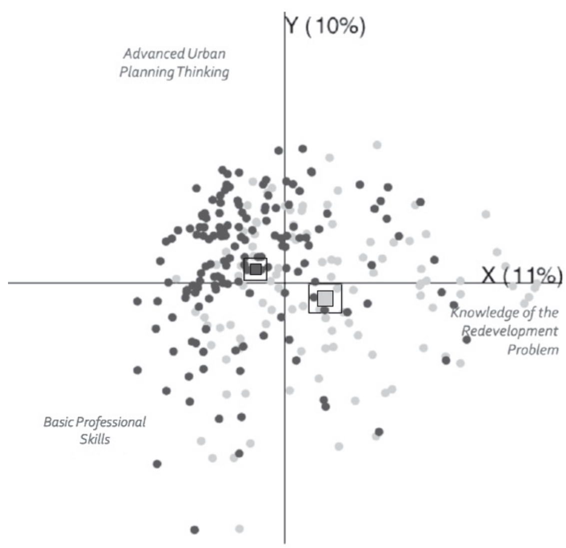


Figure 15.3. Centroids of college students (dark) and high school students (light) with the corresponding means (squares) and confidence intervals (boxes).

[1] It is possible, of course, that two networks with very different structures of connections will share similar centroids. For example, a network with many connections might have a centroid near the origin; but the same would be true of a network that had only a few connections at the far right and a few at the far left of the network space. For obvious reasons, no summary statistic in a dimensional reduction can preserve all of the information of the original network.

they are more similar to the high school students. The mean position of the gifted high school students in the projection space (Figure 15.4) is statistically significantly different from both the college students and the high school students only on the first (x) dimension:

$\bar{x}_{GiftedHS} = 0.007$, $\bar{x}_{College} = -0.083$, $t = 2.538$, $p = 0.013$, *Cohen's* $d = 0.202$

$\bar{x}_{GiftedHS} = 0.007$, $\bar{x}_{College} = 0.115$, $t = -2.736$, $p = 0.007$, *Cohen's* $d = -0.223$

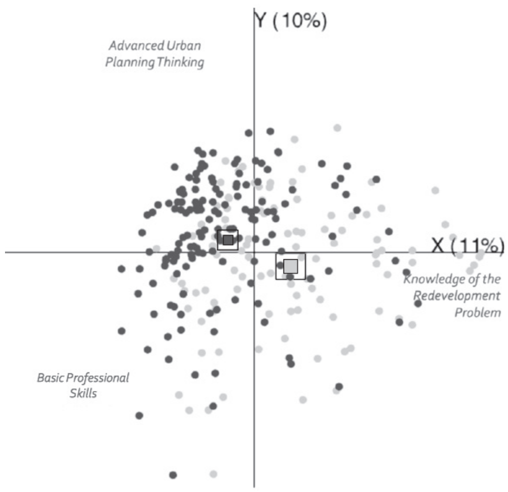


Figure 15.4. Mean network positions (squares) and confidence intervals (boxes) of the college students (left), high school students (right), and gifted high school students (center).

To determine what factors account for the differences among the three groups, we can compare their mean epistemic networks. As Figure 15.5 shows, the gifted high school students on average made more and stronger connections to elements of advanced urban planning thinking than the high school students, but not to the same extent as the college students. That is, they were somewhere between the high school and college students with respect to complex thinking in the domain. In contrast, the gifted high school students seem to be more similar to the high school students in that both populations made fewer connections than the college students between basic professional skills and advanced urban planning thinking. In other words, the gifted high school students are somewhere between the high school and college students intellectually, but they are more similar to the high school students in their level of basic professional and interpersonal skills.

Qualitative Triangulation of ENA Network Models

A key feature of ENA is the ability to trace connections in the model back to the original data – the chats, in this case – on which the connections are based. By clicking on the line connecting "epistemology of social issues" with "knowledge of data," we can access all the utterances that contributed to this connection in the network graph. Figure 15.6 shows an excerpt of the utterances that contributed to this connection in one college student's epistemic network.

The text is coloured such that stanzas or utterances containing only the first code are shown in red, those containing only the second code are shown in blue, those containing both codes are shown in purple, and those containing neither code are shown in black. The stanza (i.e., the activity) "Final Proposal Reflection," for example, is coloured purple because it contains utterances coded for both E.social.issues and K.data: the first (red) utterance justifies a land-use change based on a desire to improve the city (epistemology of social issues), while the second utterance references knowledge about the effects of zoning changes on atmospheric carbon dioxide levels (knowledge of data).

This feature of ENA allows us to close the interpretive loop (see Figure 15.7). We started with a dataset that was coded for urban planning epistemic frame elements; we used the coded data to create and visualize network models of students' urban planning thinking based on the co-occurrence of frame elements; then, if we want to understand the basis for any of the connections in the network models, we can return to the original utterances. ENA thus enables quantitative analysis of qualitative data in such a way that the quantitative results can be validated qualitatively.

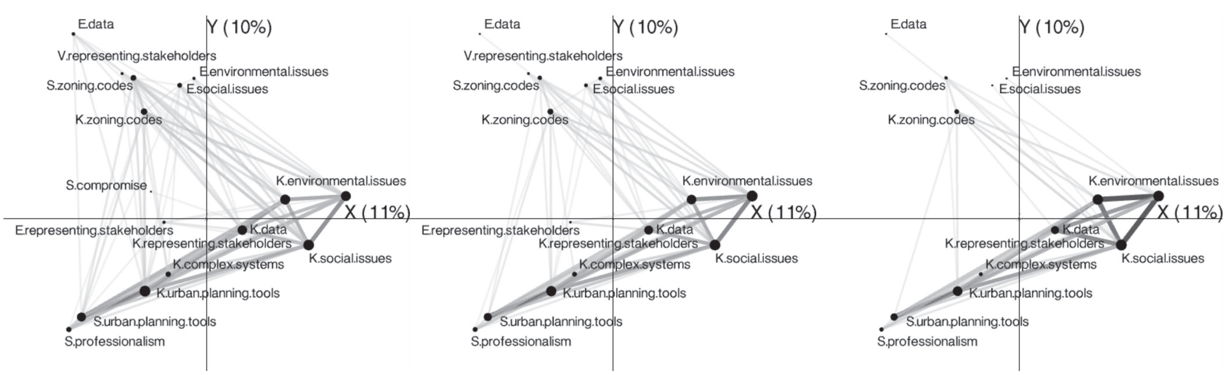
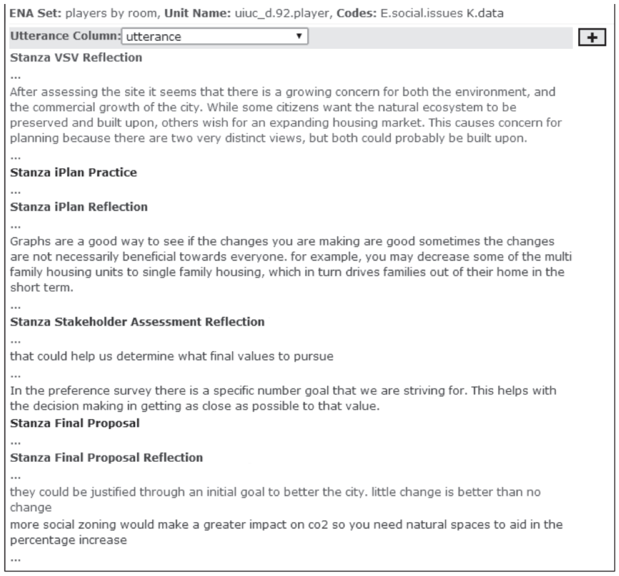


Figure 15.5. Mean epistemic networks of college students (red, left), gifted high school students (green, centre) and high school students (blue, right).



ENA Set: players by room, Unit Name: uiuc_d.92.player, Codes: E.social.issues K.data

Utterance Column: utterance

Stanza VSV Reflection

...

After assessing the site it seems that there is a growing concern for both the environment, and the commercial growth of the city. While some citizens want the natural ecosystem to be preserved and built upon, others wish for an expanding housing market. This causes concern for planning because there are two very distinct views, but both could probably be built upon.

...

Stanza iPlan Practice

...

Stanza iPlan Reflection

...

Graphs are a good way to see if the changes you are making are good sometimes the changes are not necessarily beneficial towards everyone. for example, you may decrease some of the multi family housing units to single family housing, which in turn drives families out of their home in the short term.

...

Stanza Stakeholder Assessment Reflection

...

that could help us determine what final values to pursue

...

In the preference survey there is a specific number goal that we are striving for. This helps with the decision making in getting as close as possible to that value.

Stanza Final Proposal

...

Stanza Final Proposal Reflection

...

they could be justified through an initial goal to better the city. little change is better than no change

more social zoning would make a greater impact on co2 so you need natural spaces to aid in the percentage increase

...

Figure 15.6. Excerpt of the chat utterances that contributed to the connection between epistemology of social issues and knowledge of data in one college student's epistemic network. In the ENA toolkit, the text of each utterance is coloured to indicate whether it contains code A (red), code B (blue), both (purple), or neither (black).

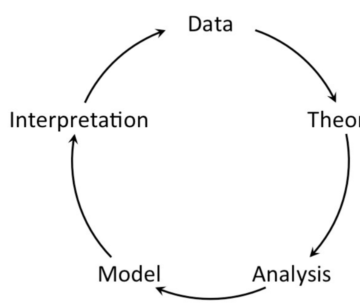


Figure 15.7. Good theory-based learning analytics "closes the interpretive loop" by making it possible to validate the interpretation of a model against the original data.

DISCUSSION

In working through this analysis, our aim was not to provide an *ideal* example for others to emulate, nor to suggest that epistemic frame theory has any particular analytic advantages over other learning theories, but to provide context for a more general discussion of methodology in learning analytics and educational data mining. As analyses of large educational datasets have become more common, a key application is obtaining empirical evidence to "refine and extend educational theories and well-known educational phenomena, towards gaining deeper understanding of the key factors impacting learning" (Baker & Yacef, 2009, p. 7). In other words, a theoretical framework guides the selection of variables and development of hypotheses, which can lead to an explanation for why observed phenomena are occurring.

In the worked example presented above, we used the theory of epistemic frames to guide our analysis of student chat data in an urban planning simulation. Epistemic frame theory suggests that learning can be characterized by the structure of connections that students make among elements of authentic practice. Our analytic approach, ENA, uses discourse data to construct models of student learning that are visualized as network graphs, mathematical representations of patterns of connections. The analysis is thus an operationalization of a particular theoretical approach to understanding learning.

One way to conceptualize the linkage between theory, data, and analysis is through *evidence centred design* (Mislevy & Riconscente, 2006; Rupp, Gustha et al., 2010; Shaffer et al., 2009). In evidence-centred design, an analytic framework is composed of three connected models: a student model, an evidence model, and a task model (see Figure 8; Mislevy, Steinberg, & Almond, 1999; Mislevy, 2006). The *student model* represents the characteristics of the student that we want to assess, or more generally the outcome we are trying to model or measure. The *task model* represents the activities and the data that will be used to measure the outcomes in the student model. The student (outcome) model and task (data) model are linked by an *evidence model*, which details the analytic tools and techniques that will be used to warrant conclusions about the outcomes based on the data.

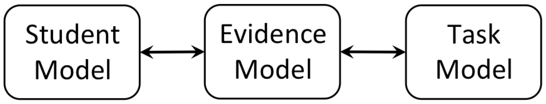


Figure 15.9. Models in an ECD Assessment (adapted from Mislevy, 2006).

Our worked example illustrates an approach to learning analytics in which each of the models (student, evidence, and task) are derived from the same theoretical framework – in this case, epistemic frame theory (see Figure 15.9).

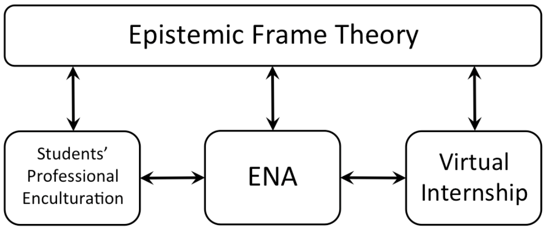


Figure 15.9. Mean network positions (squares) and confidence intervals (boxes) of the college students (left), high school students (right), and gifted high school students (center).

The result is an approach to analyzing expertise in the context of (simulated) complex problem solving that is guided by a particular theory of expertise and validated empirically. But critically, the empirical grounding of the results does not rely solely on statistical significance: because of the linkages between the different models or layers of the evidentiary argument, the interpretation of the statistics – the meaning of the model – can be verified in the original data.

Despite these advantages of a theory-based approach to data analysis, there has been a significant expansion in studies that take a radically atheoretical approach to discovery. Wired editor-in-chief Chris Anderson (2008) has even claimed that theory-based inquiry is unnecessary in the age of big data. "Petabytes [of data] allow us to say: 'Correlation is enough,'" Anderson suggests. "We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot." Despite the fact that most scientists would be deeply uncomfortable with the idea that causation is unimportant, Anderson's approach to the analysis of big data – "to view data mathematically first and establish a context for it later" – is a commonly applied method in data mining.

Of course, with a sufficiently large dataset and the ability to run it through dozens if not hundreds of mathematical models, statistically significant patterns will be found. But statistical significance does not imply conceptual or even practical significance. This does not imply that all theory-based approaches to analyzing large collections of data are ideal or even worthwhile. There is bad theory, just as there is bad empiricism – and even good theory badly operationalized or applied. Nor are we suggesting that the worked example above, or even more generally the theories and methods that we chose, are ideal in all circumstances.

Our argument, rather, is that there are distinct advantages to taking a theory-based approach to the analysis of large educational datasets. The worked example above illustrates how in theory-guided learning analytics, an explicit theoretical framework guides the search for understanding in a corpus of data and the selection of appropriate analytic methods. These linkages between data, theory, and analysis thus provide the ability to interpret the results sensibly and meaningfully.

ACKNOWLEDGEMENTS

This work was funded in part by the National Science Foundation (DRL-0918409, DRL-0946372, DRL-1247262, DRL-1418288, DUE-0919347, DUE-1225885, EEC-1232656, EEC-1340402, REC-0347000), the MacArthur Foundation, the Spencer Foundation, the Wisconsin Alumni Research Foundation, and the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison. The opinions, findings, and conclusions do not reflect the views of the funding agencies, co-operating institutions, or other individuals.

REFERENCES

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired*, *16*(7). https://www.wired.com/2008/06/pb-theory/

Andrist, S., Collier, W., Gleicher, M., Mutlu, B., & Shaffer, D. (2015). Look together: Analyzing gaze coordination with epistemic network analysis. *Frontiers in Psychology*, 6(1016). journal.frontiersin.org/article/10.3389/fpsyg.2015.01016/pdf

Arastoopour, G., Shaffer, D. W., Swiecki, Z., Ruis, A. R., & Chesler, N. C. (2016). Teaching and assessing engineering design thinking with virtual internships and epistemic network analysis. *International Journal of Engineering Education*, 32(2), in press.

Atkinson, R., Derry, S., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research*, 70(2), 181–214.

Bagley, E. A., & Shaffer, D. W. (2009). When people get in the way: Promoting civic thinking through epistemic game play. *International Journal of Gaming and Computer-Mediated Simulations*, 1(1), 36–52.

Bagley, E. A., & Shaffer, D. W. (2015a). Learning in an urban and regional planning practicum: The view from educational ethnography. *Journal of Interactive Learning Research*, 26(4), 369–393.

Bagley, E. A., & Shaffer, D. W. (2015b). Stop talking and type: Comparing virtual and face-to-face mentoring in an epistemic game. *Journal of Computer Assisted Learning*, 31(6), 606–622.

Baker, R. S. J. d., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17.

Chesler, N. C., Ruis, A. R., Collier, W., Swiecki, Z., Arastoopour, G., & Shaffer, D. W. (2015). A novel paradigm for engineering education: Virtual internships with individualized mentoring and assessment of engineering thinking. *Journal of Biomechanical Engineering*, 137(2), 024701:1–8.

Dorogovtsev, S. N., & Mendes, J. F. F. (2013). *Evolution of networks: From biological nets to the Internet and WWW*. Oxford, UK: Oxford University Press.

Gee, J. P. (1999). *An introduction to discourse analysis: Theory and method*. London: Routledge.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Cambridge, MA: Harvard University Press.

Goodwin, C. (1994). Professional vision. *American Anthropologist*, 96(3), 606–633.

i Cancho, R. F., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London B: Biological Sciences*, 268(1482), 2261–2265.

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Erlbaum.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, MA: Cambridge University Press.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instruments, & Computers*, 28(2), 203–208.

Mislevy, R. J. (2006). *Issues of structure and issues of scale in assessment from a situative/sociocultural perspective*. CSE Technical Report 668. Los Angeles, CA.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (1999). *Evidence-centered assessment design*. Educational Testing Service. http://www.education.umd.edu/EDMS/mislevy/papers/ECD_overview.html

Mislevy, R., & Riconscente, M. M. (2006). Evidence-centered assessment design: Layers, concepts, and terminology. In S. Downing & T. Haladyna (Eds.), *Handbook of Test Development* (pp. 61–90). Mahwah, NJ: Erlbaum.

Nash, P., Bagley, E. A., & Shaffer, D. W. (2012). Playing for public interest: Epistemic games as civic engagement activities. *American Educational Research Association Annual Conference* (AERA). Vancouver, BC, Canada.

Nash, P., & Shaffer, D. W. (2011). Mentor modeling: The internalization of modeled professional thinking in an epistemic game. *Journal of Computer Assisted Learning*, 27(2), 173–189.

Nash, P., & Shaffer, D. W. (2012). Epistemic youth development: Educational games as youth development activities. *American Educational Research Association Annual Conference* (AERA). Vancouver, BC, Canada.

Nash, P., & Shaffer, D. W. (2013). Epistemic trajectories: Mentoring in a game design practicum. *Instructional Science*, 41(4), 745–771.

Newman, M. E. J. (2004). Analysis of weighted networks. *Physical Review E*, 70(5), 56131.

Rohde, M., & Shaffer, D. W. (2004). Us, ourselves and we: Thoughts about social (self-) categorization. *Association for Computing Machinery (ACM) SigGROUP Bulletin*, 24(3), 19–24.

Rupp, A. A., Sweet, S., & Choi, Y. (2010). Modeling learning trajectories with epistemic network analysis: A simulation-based investigation of a novel analytic method for epistemic games. In R. S. J. d. Baker, A. Merceron, P. I. Pavlik Jr. (Eds.), *Proceedings of the 3rd International Conference on Educational Data Mining* (EDM2010), 11–13 June 2010, Pittsburgh, PA, USA (pp. 319–320). International Educational Data Mining Society.

Shaffer, D. W. (2004). Epistemic frames and islands of expertise: Learning from infusion experiences. *Proceedings of the 6th International Conference of the Learning Sciences* (ICLS 2004): *Embracing Diversity in the Learning Sciences*, 22–26 June 2004, Santa Monica, CA, USA (pp. 473–480). Mahwah, NJ: Lawrence Erlbaum.

Shaffer, D. W. (2006). Epistemic frames for epistemic games. *Computers and Education*, 46(3), 223–234.

Shaffer, D. W. (2007). *How computer games help children learn*. New York: Palgrave Macmillan.

Shaffer, D. W. (2012). Models of situated action: Computer games and the problem of transfer. In C. Steinkuehler, K. D. Squire, & S. A. Barab (Eds.), *Games, learning, and society: Learning and meaning in the digital age* (pp. 403–431). Cambridge, UK: Cambridge University Press.

Shaffer, D. W., Collier, W., & Ruis, A. R. (2016). A tutorial on epistemic network analysis: Analyzing the structure of connections in cognitive, social, and interaction data. *Journal of Learning Analytics*, 3(3), 9–45.

Shaffer, D. W., Hatfield, D. L., Svarovsky, G. N., Nash, P., Nulty, A., Bagley, E. A., … Frank, K. (2009). Epistemic Network Analysis: A prototype for 21st century assessment of learning. *International Journal of Learning and Media*, 1(1), 1–21.

Svarovsky, G. N. (2011). Exploring complex engineering learning over time with epistemic network analysis. *Journal of Pre-College Engineering Education Research*, 1(2), 19–30.

Tannen, D. (1993). *Framing in discourse*. Oxford, UK: Oxford University Press.

Wise, A. F., & Shaffer, D. W. (2016). Why theory matters more than ever in the age of big data. Journal of Learning Analytics, 2(2), 5–13.

APPENDIX I

URBAN PLANNING EPISTEMIC FRAME CODE SET

Code	Code Description	Example
Epistemology of Social Issues	Using social issues to justify a decision or to ask for a justification of a decision (e.g., jobs, crime, housing)	Because it effects their [the stakeholders'] business
Epistemology of Environmental Issues	Using environmental issues to justify a decision or to ask for a justification of a decision (e.g., runoff, pollution, animal habitats)	except why wouldn't we think that they'd care about social and environmental issues?
Epistemology of Representing Stakeholders	Using the representation of stakeholders to justify a decision or to ask for a justification of a decision (e.g., referring to a specific stakeholders' needs by name, referring to the needs of the stakeholder group)	Try and understand what the stakeholders want and why. That might help you come up with a plan that they'll support.
Epistemology of Data	Using data to justify a decision or to ask for a justification of a decision (e.g., numbers, collecting information)	there are three different groups L-EDC, L-CAG, and LC-RWC and each group have different recommendations. all three of the numbers are different so it seems impossible to meet all three of those numbers at the same time. which group are we suppose to follow?
Epistemology of Compromise	Using compromise to justify a decision or to ask for a justification of a decision (e.g., balancing stakeholders' needs, referring explicitly to compromise)	You may have to make compromises, because the stakeholder groups sometimes disagree.
Value of Representing Stakeholders	Utterances indicating that players should, should not, must, must not, ought to care about representing stakeholders	First, let's make sure we agree on what the stakeholders want.
Value of Complex Systems	Utterances indicating that players should, should not, must, must not, ought to care about relationships between parts of a larger system	Flavian and Natalie both want to protect the enviornment in Lowell, whereas Lee and Nathanial both want to increase housing and economic groth
Value of Compromise	Utterances indicating that players should, should not, must, must not, ought to care about compromise	however, we may need to make compromises so everyone can live with the changes.
Skill of Professionalism	Utterance indicating that a skill related to professionalism was performed (e.g., sending an email)	OK. I finished the interview and I read the resources.
Skill of Data	Utterance indicating that a skill related to data was performed (e.g., entering values into the TIM, referring to values of TIM output and stakeholder assessment values)	we listened to the feedbacks and chose the best numbers according to what they wanted so i think 99.9999
Skill of Scientific Thinking	Utterance indicating that a skill related to scientific thinking was performed (e.g., making hypotheses, testing hypotheses, developing models)	so we put test numbers into the TIM to see how they react?
Skill of Compromise	Utterance indicating that a skill related to compromise was performed	however, we may need to make compromises so everyone can live with the changes.
Identity of Urban Planners	Utterance indicating that one or one's group identifies as an urban planner	Sure. iPlan is a model, so as a planner, when you make changes in iPlan, it shows us what might happen if you made those changes in the real world.
Identity of Interns	Utterance indicating that one or one's group identifies as interns	please remember we are professionals and all our chat and work that we hand in should reflect that.
Knowledge of Social Issues	Utterance referring to social issues (e.g., jobs, crime, housing)	I worked with a group that cared about nests, housing, phosphorous, and runoffs

Code	Code Description	Example
Knowledge of Environmental Issues	Utterance referring to environmental issues (e.g., runoff, pollution, animal habitats)	I worked with the Connecticut River Water council and they cared about the environment.
Knowledge of Representing Stakeholders	Utterance referring to representing stakeholders (e.g., referring to a specific stakeholders' needs by name, referring to the needs of the stakeholder group)	You may have to make compromises, because the stakeholder groups sometimes disagree
Knowledge of Complex Systems	Utterance referring to relationships between parts of a larger system	inorder to reduce co2 levels I had to increase the bird populations
Knowledge of Urban Planning Tools	Utterance referring to urban planning tools (e.g., iPlan, TIM, Preference Survey)	so wer making one iplan?
Knowledge of Zoning Codes	Utterance referring to zoning codes (e.g., open space, industrial space, housing, wetlands)	well if you changed a piece of land from open space to industrial, you create jobs, but the CO might increase as well
Knowledge of Data	Utterance referring to data (e.g., entering values into the TIM, referring to values of TIM output and stakeholder assessment values)	i got really positive feedback except for my runoff number. i need to reduce that a bit more
Knowledge of Scientific Thinking	Utterance referring to scientific thinking (e.g., making hypothesis, testing hypothesis, developing models)	I guess it would be a model that you can test and observe the effects that would happen in the real world.
Skill of Zoning Codes	Utterance indicating that a skill related to zoning codes was performed	Converting land to open space/wetlands increases the bird population. That's one of the stakeholder's concerns, so I should maybe mark more open space and less industrial or commercial
Skill of Urban Planning Tools	Utterance indicating that a skill related to urban planning tools was performed	To get the desired result you'd have to change a few different indicators.

Chapter 16: Multilevel Analysis of Activity and Actors in Heterogeneous Networked Learning Environments

Daniel D. Suthers

Department of Information and Computer Sciences, University of Hawaii, USA

DOI: 10.18608/hla17.016

ABSTRACT

Learning in today's networked environments is often distributed across multiple media and sites, and takes place simultaneously via multiple levels of agency and processes. This is a challenge for those wishing to study learning as embedded in social networks, or simply to monitor a networked learning environment for practical purposes. Traces of activity may be fragmented across multiple logs, and the granularity at which events are recorded may not match analytic needs. This chapter describes an analytic framework, Traces, for analyzing participant interaction in one or more digital settings by computationally deriving higher levels of description. The Traces framework includes concepts for modelling interaction in sociotechnical systems, a hierarchy of models with corresponding representations, and computational methods for translating between these levels by transforming representations. Potential applications include identifying sessions of interaction, key actors within sessions, relationships between actors, changes in participation over time, and groups or communities of learners.

Keywords: Interaction analysis, social network analysis, community detection, networked learning environments, Traces analytic framework

This chapter is most relevant to readers who will be analyzing trace data of participant interaction in one or more digital settings (which may be of multiple media types), and wish to computationally derive higher levels of description of what is happening, possibly across multiple settings. Examples of these higher levels of description include identifying sessions of interaction, identifying groups or communities of learners across sessions, characterizing interaction in terms of key actors, and identifying relationships between actors. The approach outlined here, the Traces framework, has been used for discovery oriented research, but can also support hypothesis testing research that requires variables at these higher levels of description, or live monitoring of production learning settings using such descriptions. The *Traces framework* involves a set of concepts for thinking about and modelling interaction in sociotechnical systems, a hierarchy of models with corresponding representations, and computational methods for translating between these levels by transforming representations. These methods have been implemented in experimental software and tested on data from a heterogeneous networked learning environment.

The purpose of this chapter is to introduce the reader to the conceptual and representational aspects of the framework, with brief descriptions of how it can be used for multilevel analysis of activity and actors in networked learning environments. Due to length limitations, detailed examples and information on our implementation and research are not included[1].

[1] See Joseph, Lid, and Suthers (2007) and Suthers (2006) for theoretical background; see Suthers, Dwyer, Medina, and Vatrapu (2010) and Suthers and Rosen (2011) for the development of our analytic representations; see Suthers, Fusco, Schank, Chu, and Schlager (2013) for community detection applications; see Suthers (2015) for an example of how one might combine these capabilities into an activity reporter for monitoring a large networked learning environment. Suthers et al. (2013) and Suthers (2015) describe the data from the Tapped In network of educators we used as a case study in developing this framework. Papers are available at http://lilt.ics.hawaii.edu.

MOTIVATIONS

Motivations for the Traces framework derive in part from phenomena such as the emergence of Web 2.0 (O'Reilly, 2005) and its adoption by educational practitioners and learners for formal and informal learning, including more recent interest in MOOCs (massive open online courses) (Allen & Seaman, 2013). In these environments, learning is distributed across time and virtual place (media), and learners may participate in multiple settings. We focus on networked learning environments (NLE), which we define to include any sociotechnical network that involves mediated interaction between participants (hence "networked") in which learning might take place, including for example online communities (Barab, Kling, & Gray, 2004; Renninger & Shumar, 2002) and cMOOCs (connectivist MOOCs) (Siemens, 2013). The framework is not applicable to isolated activity by individuals, such as "xMOOCs" in which large numbers of individuals interact primarily with courseware or tutoring systems.

Learning and knowledge creation activities in these networked environments are often distributed across multiple media and sites. As a result, traces of such activity may be fragmented across multiple logs. For example, networked learning environments may include a mashup of threaded discussion, synchronous chats, wikis, microblogging, whiteboards, profiles, and resource sharing. Events may be logged in different formats and locations, disassociating actions that for participants were part of a single unified activity. Integration of multiple sources of trace data into a single transcript may be needed to reassemble data on the interaction. Also, the granularity at which events are recorded may not match analytic needs. For example, media-level events may be the wrong ontology for analyses concerned with relationships between acts, persons, and/or media rather than individual acts. Translation from log file representations to other levels of description may be required to begin the primary analysis.

Derivation of higher levels of description is also motivated by theoretical accounts of learning as a complex and multilevel phenomenon. Theories of how learning takes place in social settings vary regarding the agent of learning, including individual, small group, network, or community; and in the process of learning, including for example information transfer, argumentation, intersubjective meaning-making, shifts in participation and identity, and accretion of cultural capital (Suthers, 2006). Learning takes place simultaneously at all of these levels of agency and with all of these processes, potentially at multiple time scales (Lemke, 2000). A multi-level approach is also motivated by our theoretical stance that social regularities arise from how myriad individual acts are aggregated and influence each other, possibly mediated by artifacts (Latour, 2005), and the methodological implication that to understand phenomena such as actor relationships or community structures, we also need to look at the stream of individual acts out of which these phenomena are constructed. Thus, understanding learning in its full richness requires data that reveal the relationships between individual and collective levels of agency and potentially coordinating multiple theories and methods of analysis (de Laat, 2006; Monge & Contractor, 2003; Suthers, Lund, Rosé, Teplovs, & Law, 2013).

TRACES ANALYTIC FRAMEWORK

This section covers the levels of description and corresponding representations underlying the Traces analytic framework with the next section discussing potential applications. To preview the approach, logs of events are abstracted and merged into a single abstract transcript of events, which is then used to derive a series of representations that support levels of analysis of interaction and of relationships. Three kinds of graphs model interaction. *Contingency graphs* record how events such as chatting or posting a message are observably related to prior events by temporal and spatial proximity and by content. *Uptake graphs* aggregate the multiple contingencies between each pair of events to model how each given act may be "taking up" prior acts. *Session graphs* are abstractions of uptake graphs: they cluster events into spatio-temporal sessions with uptake relationships between sessions. Relationships between actors and artifacts are abstracted from interaction graphs to obtain *sociograms* and a special kind of affiliation network that we call *associograms*. The representations used at various levels of analysis are shown schematically in Figure 16.1.

About Transcript

We begin with various traces of activity (such as log files of events) that provide the source data (Figure 16.1a). These are parsed, using necessarily system-specific methods, into an event stream, as shown in the second level (boxes in Figure 16.1b). Events can come from different media (e.g., chats, threaded discussion, social media), be of various types (e.g., enter chat, exit chat, chat contribution, post message, read message, download file). They should be annotated with time stamps, actors, content (e.g., chat content), and locations (e.g., chat rooms) involved in the event where relevant. The result is an abstract transcript of the distributed activities. By translating from system-specific representations of activity to the abstract transcript, we integrate hitherto fragmented records of activity into one analytic artifact.

Contingency Graph

We then compute *contingencies* between events (arrows in Figure 16.1b), to produce a model of how acts are mutually contextualized. Human action is contingent upon its setting in diverse ways: computational methods can capture some of the contingencies amenable to automated detection. For example, a contingency called *proximal event* reflects the likelihood that events occurring close together in time and space are related. In analyzing quasi-synchronous chat, contingencies are installed to prior contributions in the same room that occur within an adjustable time window but not too recently. *Address* and *reply* contingencies are installed between an utterance mentioning a user by name and the last contribution and next contribution by that participant within a time window, using a parser/matcher of user IDs to first names. *Same actor* contingencies are installed to prior acts of a participant over a larger time window to reflect the continuity of an agent's purpose. Overlap in content as represented by sets of lexical stems is used to produce a *lexical overlap* contingency weighted by the number of overlapping stems. Further contingencies could be computed based on natural language processing methods for analysis of interactional structure (Rosé et al., 2008).

The resulting contingency graph represents the first layer of abstraction (Figure 16.1b), the contextualized action model. In this graph, vertices are events, and contingencies are typed edges between vertices (example types were just described). There may be multiple edges between any two vertices (e.g., two proximal events by the same actor with lexical overlap will have at least three contingencies between them).

Uptake Graph

It is necessary to collapse the multiple edges between vertices into single edges for two reasons. First, most graph algorithms assume at most only one edge between any two vertices. Second, we are interested in *uptake*, the relationship between events in which a human action takes up some aspects of prior events as being significant in some manner. Being the fundamental building block of interaction (Suthers et al., 2010), uptake is a basic unit for analysis of how learning takes place in and through interaction. Replying to prior contributions in chats and discussions are examples of uptake, but uptake is not limited to replies: one can appropriate a prior actor's contribution in other ways. Uptake is not specific to a medium: it can occur in different media, and cross media (Suthers et al., 2010). Contingencies are of interest only as collective evidence for uptake, so we abstract the contingency

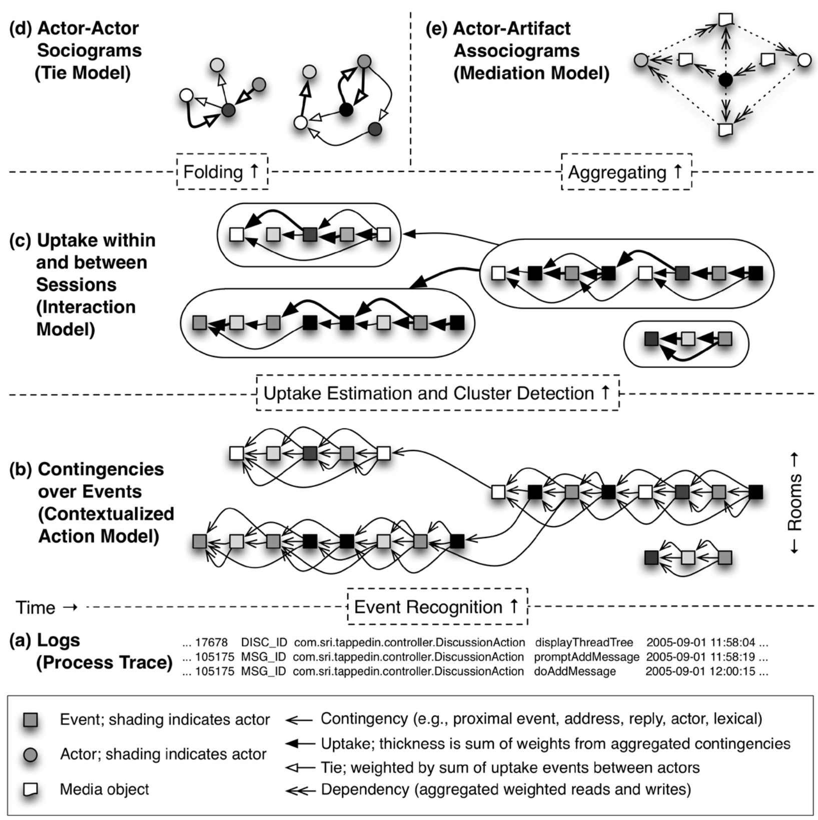


Figure 16.1. Levels of analysis and their representations.

graph to an uptake graph.

As shown in Figure 16.1c, uptake graphs are similar to contingency graphs in that they also relate events, but they collect together bundles of the various types of contingencies between a given pair of vertices into a single graph edge, weighted by a combination of the strength of evidence in the contingencies and optionally filtering out low-weighted bundles (Suthers, 2015). Different weights can be used for different purposes (e.g., finding sessions, analyzing the interactional structure of sessions, constructing sociograms). Importantly, we do not throw away the contingency weights: these are retained in a vector to summarize the nature of the uptake relation, and, once aggregated into sociograms, of the tie between actors. We can do several interesting things with uptake graphs, but first we usually identify portions of the graph that we want to handle separately, as they represent sessions.

Sessions

Clusters of events in spatio-temporal proximity are computed to identify sessions (indicated by rounded containers in Figure 16.1c). Methods for doing so are discussed later. For intra-session analysis, the uptake graph for a session is isolated. Several paths are possible from here. For example, the sequential structure of the interaction can be micro-analyzed to understand the development of group accomplishments: this analysis may be difficult to automate. Methods for graph structure analysis can be applied, such as cluster detection, or tracing out thematic threads (Trausan-Matu & Rebedea, 2010). For inter-session analysis, we collapse each session into a single vertex representing the session, but retain the inter-session uptake links. (For example, there are four sessions in Figure 16.1c and two inter-session uptakes.) These inter-session links indicate potential influences across time and space from one session to another.

Sociograms

Sociotechnical networks are commonly studied using the methods of social network analysis, using *sociogram* or *sociomatrix* representations of the presence or strength of ties between human actors, and graph algorithms that leverage the power of these representations to expose both local (ego-centric) and non-local (network) social structures (Newman, 2010; Wasserman & Faust, 1994). Either within or across sessions, we can fold uptake graphs into actor–actor sociograms (directed weighted graphs, Figure 16.1d). The tie strength between actors is the sum of the strength of uptake between their contributions. If we want to be stricter about the evidence for relations between the two actors, we may use a different weighting that downplays proximity and emphasizes direct evidence of orientation to the other actor. These sociograms can be analyzed using conventional social network analysis methods, for example centrality metrics to identify key actors.

Associograms

The sociogram's singular tie between two actors summarizes yet obscures the many interactions between the actors on which the tie is based, as well as the media through which they interacted. To retain the advantages of graph computations on a summary representation while retaining some of the information about how the actors interacted, we use bipartite, multimodal, directed weighted graphs, similar to but more specific than affiliation networks. They are bipartite because all edges go strictly between actors and artifacts and multimodal because the artifact nodes can be categorized into the different kinds of mediators that they represent; for example, chat rooms, discussion forums, and files. Directed edges (arcs) indicate read/write relations or their analogs: an arc goes from an actor to an artifact if the actor has read that artifact (e.g., opened a discussion message or was present when someone chatted), and from an artifact to an actor if the actor modified the artifact (e.g., posted a discussion message or chatted). The direction of the arc indicates a form of dependency, the reverse direction of information flow. Weights on the arcs indicate the number of events that took place between the corresponding actor/artifact pair in the indicated direction. Since "affiliation network" is not specific enough and "bipartite multimodal directed weighted graph" is too long, to highlight their unique nature we call these graphs *associograms* (Suthers & Rosen, 2011). This term is inspired by Latour's (2005) concept that social phenomena emerge from dynamic networks of associations between human and non-human actors.

For example, Figure 16.2 shows a portion of an associogram from the Tapped In educator network,

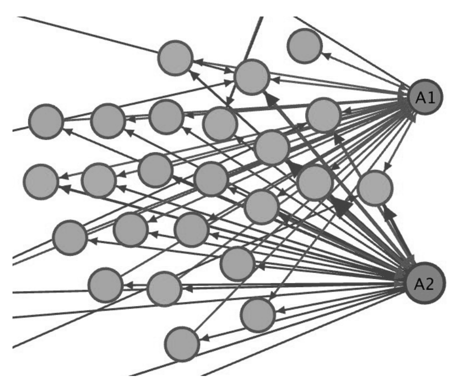


Figure 16.2. An associogram from the Tapped In data. Actors represented by nodes on the right have read and written to the files and discussions represented by differently coloured nodes on the left.

representing asymmetric interaction between two actors, with one actor writing most of the files and another writing to most of the discussions. A sociogram consisting of a single link between these two actors would fail to capture this information. The associogram retains information about the distribution of activity across media. Network analytic methods can then simultaneously tell us how both human actors and artifacts participate in generating the larger phenomena of interest, such as the presence of communities of actors and the media through which they are technologically embedded (Licoppe & Smoreda, 2005). Although interaction is not directly represented, the associogram also provides a bridge to the interaction level of analysis (Suthers et al., 2010), allowing us to retrieve activity in specific media settings.

EXAMPLES OF ANALYTIC OPTIONS

The Traces framework provides multiple pathways for analysis. In the following sections we illustrate various analyses that can be supported by this framework (Suthers & Dwyer, 2015). These examples are from analyses we have done with our experimental software implementation.

Identifying Sessions of Interaction

Different options exist for detection of sessions in interaction graphs. If interaction is not clearly demarcated by periods of non-interaction and one wishes to discover clusters of high activity, we have found that cohesive subgraph detection or "community detection" algorithms (Fortunato, 2010) such as modularity partitioning (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) applied to uptake graphs are useful (Suthers, 2017). If (as in our Tapped In data) activity is distributed across rooms and the activity within a room almost always has periods of non-activity between sessions, sessions can be identified efficiently without needing to

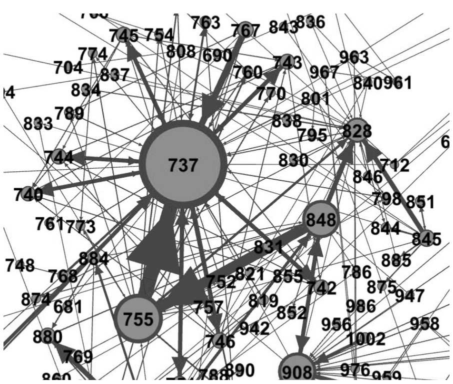


Figure 16.3. Close-up of session graph with inter-session uptake. Node and edge size is weighted degree.

construct a contingency graph (it can be constructed later for other purposes). Activity is tracked in each room, and a new session ID is assigned to the room every time there is a gap of S seconds of no activity. S is a tunable parameter, such as 240 seconds. Suthers (2017) discusses these options further.

Tracing Influences Between Sessions

We might be interested in non-local influences between sessions across time and space. Uptake relations between events in different sessions can be aggregated into weighted uptake relations between sessions (Figure 16.1c). An example *session graph* from Suthers (2015) is shown in Figure 16.3. Some sessions have more heavily weighted links between them. Reading the edges in reverse order (uptake points backwards in time), we see that session 737 (in the Reception room) influenced session 755 (Teaching Teachers room), which in turn influenced 848 (NTraining room). Examination of the rooms and participants involved showed that many participants logged into or met in the Reception room, then went to Teaching Teachers for session 755 on mentoring in the schools. Then the facilitator of 755 announced that she had another session on teacher training in another room: several participants in the mentoring session followed her to NTraining for session 848. Further details are in Suthers (2015).

Identifying Actor Roles and Tracking Change in Participation Over Time

Educators or NLE facilitators may want to identify the key participants in their online learning communities, whether for assessment in formal educational settings, to encourage volunteers in participant-driven settings, or for research purposes such as to study what drives key participants. It is also important to know who is disengaged. Some of these needs can be met through social network analysis. We can generate sociograms for any granularity of the uptake graph (e.g., within a session, or across sessions over a time period) by folding uptake relations between events into ties between their actors. For example, a facilitator might want to see a sociogram summarizing actor activity in session 755, the session on mentoring teachers led by MT. The sociogram is shown in Figure 16.4. Node size is weighted in-degree, discussed below.

Sociograms add information over mere counts of number of contributions because some sociometrics are sensitive to the network context of nodes representing actors. For example, weighted in-degree indicates the extent to which other acts have contingencies to and hence potentially took up a given act. Aggregating these acts for an actor is an estimate of how much an actor's contributions are taken up by others. This metric is sensitive to both the level of activity of the actor and that activity's relation to others' activity. Weighted out-degree is an estimate of how much an actor takes up others' contributions. Eigenvector centrality (and its variants such as PageRank and Hubs and Authorities) is a non-local metric that takes into account the centrality of one's neighbours (Newman, 2010, p. 169), indicating the extent to which an actor is connected to others who are themselves central. Betweeness centrality is an indicator of actors who potentially play brokerage roles in the network: high betweeness centrality means that the node representing an actor is on relatively more shortest paths between other actors (Newman, 2010, p. 185), so potentially controls information flow or mediates contact between these actors. Betweeness is of particular interest when examining activity across sessions: different sessions generally have different actors, so an actor attending multiple sessions will have high betweeness.

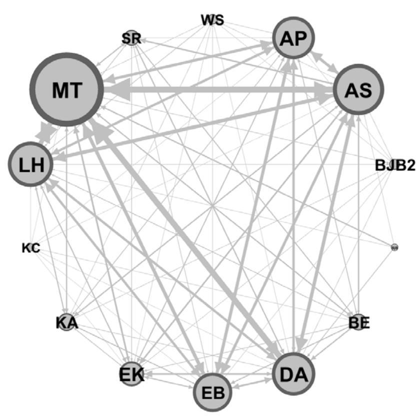


Figure 16.4. Sociogram from a Tapped In session showing prominent actors and their interactions.

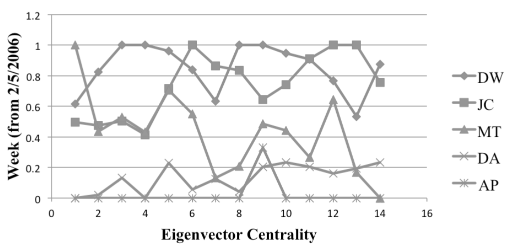


Figure 16.5. Eigenvector Centrality of several actors in Tapped In over a 14 week period.

Analyses on longer time scales may be of interest to researchers as well as practicing educators. One can trace the development of actors' roles over time in terms of changes in their sociometrics. For example, one might aggregate uptake for all actors in the network into sociograms at one week intervals, and then graph the sociometrics on a weekly basis, looking for trends. One can see some of these trends in Figure 16.5, taken from Suthers (2015). The plots for sustaining actors will remain high (e.g., DW, a volunteer guide), and for those who return for periodic events will exhibit a regular spiking structure (e.g., MT, who facilitated monthly events). Steadily increasing or decreasing metrics indicate persons becoming more or less engaged, respectively (possibly DA), and single spikes indicate a one-off visit (AP).

Superficially, these analyses appear similar to the many sociometric analyses found in the literature, so we should highlight what the Traces framework has added. Our implementation of the Traces framework derived these *latent* ties from *automated* interaction analysis of streams of events, by identifying and then aggregating multiple contingencies between events, and then folding the resulting uptake relations between events into an actor–actor graph. This has significant advantages over, for example, manual content analysis or the use of surveys to derive tie data, which are labour intensive, or reliance on explicit "friending" relations: surveys and friend links may not reflect the latent ties in actual interaction between the persons in question. Another advantage is described below.

Identifying Relationships Between Actors

The Traces framework derives ties between actors by aggregating multiple contingencies between their contributions. The contingencies indicate the qualitative nature of the relationship between these contributions, e.g., being close in time and space, using the same words, and addressing another actor by name. When contingencies are aggregated into uptake relations, we keep track of what each type of contingency contributed to the uptake relation. This record keeping is continued when folding uptakes into ties, so that for any given pair of actors we have a vector of weights that provides information about the nature of the relationship in terms of the underlying contingencies. For example, we can quantify how the relationship between DA and MT was enacted over a given time period in terms of how often they chatted in proximity to each other, the lexical overlap of their chat contents, and how often they addressed each other by name in each direction. Relational information might be of interest to educators or researchers who are managing collaborative learning activities amongst students, or even to examine one's own relations to students. The Traces framework makes this possible by retaining information about the interactional origins of ties (see Suthers, 2015).

Identifying Groups or Communities of Learners Across Sessions

In the network analysis literature, "community detection" refers to finding subgraphs of mutually associated vertices under graph-theoretic definitions, rather than to sociological concepts of community (e.g., Cohen, 1985; Tönnies, 2001). However, we can use the former as evidence for the latter, particularly when studying networked societies (Castells, 2001; Wellman et al., 2003). A good graph-theoretic definition should capture the intuition that individuals in a sociological community are more closely associated with each other than they are with individuals outside of their community. Algorithms based on the modularity metric are widely used in the literature for this purpose. The modularity metric (Newman, 2010, p. 224) compares the density of weighted links inside (non-overlapping) partitions of vertices to weighted links expected in a random graph, to find highly modular partitions. Finding the best possible partition under a modularity metric is computationally impractical on large networks, but a fast algorithm known as the Louvain method (Blondel et al., 2008) has been shown to give good approximations. An example is shown in Figure 16.6.

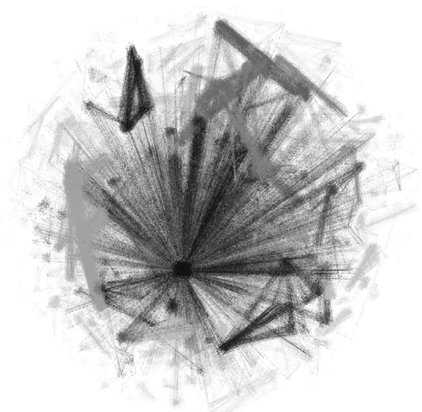


Figure 16.6. OpenOrd visualization of partitions found in combined associogram for actors associated via chats, discussions, and files in Tapped In.

Once partitions have been obtained, one can characterize the community structure of the network in several ways. Quantitative summaries can include the distribution of partition size (e.g., is there primarily one large community, or does the network contain many small and a few large communities?) and distributions of parameters across sizes (e.g., how does activity per actor relate to community size, and how does the use of different media vary with community size?). Qualitative characterization of community structure requires examining the attributes of actors and the media through which they interact to interpret each partition. See Suthers, Fusco, et al. (2013) for examples of both quantitative and qualitative analyses of the partitioning shown in Figure 16.6.

SUMMARY AND RELATED WORK

This chapter introduced the Traces analytic framework, which integrates traces of activity distributed across media, places, and time into an abstract transcript, and then provides a linked abstraction hierarchy using observable contingencies between events to build models of interaction and ties. Contingencies are applied to events in the abstract transcript to produce a contingency graph. Contingencies are then aggregated into uptake between the same events. Uptake that crosses partitions can be used to identify influences across space and time, and uptake within partitions can be analyzed to study the interactional structure of a session. Uptake graphs can be folded into networks where nodes are actors rather than events, to which sociometrics are applied. Events can also be folded into actor–artifact networks or "associograms" that capture how actors are associated with each other via mutual read and write of media objects. The framework addresses the need to understand how aggregate phenomena (e.g., "ties," "roles," and "communities") are both produced by and provide the setting of specific interactional events. It has been implemented as experimental software and tested with data from a heterogeneous networked learning community.

Other authors have noted the need to combine multiple forms of analysis, including specifically social network analysis in networked learning environments. For example, de Laat, Lally, Lipponen, and Simons (2007) and Martínez and colleagues (2006) showed the utility of combining social network analysis with various qualitative and quantitative methods in the study of participation networks. Others have constructed and folded interaction graphs into sociograms of ties between actors. For example, Rosen and Corbit (2009) constructed graphs based on temporal proximity, and Haythornthwaite and Gruzd (2008) describe preliminary work in extracting interaction relations from references and names. The Traces framework is in the same spirit, but is arguably more mature. We consider multiple kinds of relations between events to provide a richer basis for session identification and subsequent analysis of activity and actors within sessions, and have automated these analyses and tested them on a rich historical data corpus where diverse participants interacted in an environment exhibiting many features of today's distributed interaction. Work by Trausan-Matu on "polyphonic analysis" (Trausan-Matu & Rebedea, 2010) has affinities to our use of multiple contingencies, but has only recently been abstracted to higher levels of analysis. A thesis by Charles (2013) has provided an alternative implementation of our approach and extended the set of contingencies. Our approach dovetails with work that applies natural language processing methods for analysis of interactional structure, and indeed rules for generating additional contingencies could be derived from such research. Although our software is not presently in a form suitable for distribution, the author hopes that the Traces framework can serve the reader as a way to organize their own analyses of heterogeneous networked learning environments and other sociotechnical systems.

ACKNOWLEDGMENTS

Nathan Dwyer, Kar-Hai Chu, Richard Medina, Devan Rosen, and Ravi Vatrapu co-developed the ideas behind this work. Nathan Dwyer and Kar-Hai Chu also wrote the core software implementation. Mark Schlager, Patti Schank, and Judi Fusco shared data and expertise. This work was partially supported by NSF Award 0943147.

REFERENCES

Allen, I. E., & Seaman, J. (2013). *Changing course: Ten years of tracking online education in the United States.* Babson Survey Research Group. http://www.onlinelearningsurvey.com/reports/changingcourse.pdf

Barab, S. A., Kling, R., & Gray, J. H. (2004). *Designing for virtual communities in the service of learning*. New York: Cambridge University Press.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*. doi:10.1088/1742-5468/2008/10/P10008

Castells, M. (2001). *The Internet galaxy: Reflections on the Internet, business, and society*. Oxford, UK: Oxford University Press.

Charles, C. (2013). *Analysis of communication flow in online chats*. Unpublished Master's Thesis, University of Duisburg-Essen, Duisburg, Germany. (2245897)

Cohen, A. P. (1985). *The symbolic construction of community*. New York: Routledge.

de Laat, M. (2006). *Networked learning*. Apeldoorn, Netherlands: Politie Academie.

de Laat, M., Lally, V., Lipponen, L., & Simons, R.-J. (2007). Investigating patterns of interaction in networked learning and computer-supported collaborative learning: A role for social network analysis. *International Journal of Computer Supported Collaborative Learning*, *2*(1), 87–103.

Fortunato, S. (2010). *Community detection in graphs*. Physics Reports, 486, 75–174.

Haythornthwaite, C., & Gruzd, A. (2008). Analyzing networked learning texts. In V. Hodgson, C. Jones, T. Kargidis, D. McConnell, S. Retalis, D. Stamatis, & M. Zenios (Eds.), *Proceedings of the 6th International Conference on Networked Learning* (NLC 2008), 5–6 May 2008, Halkidiki, Greece (pp. 136–143).

Joseph, S., Lid, V., & Suthers, D. D. (2007). Transcendent communities. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Proceedings of the 7th International Conference on Computer-Supported Collaborative Learning* (CSCL 2007), 16–21 July 2007, New Brunswick, NJ, USA (pp. 317–319). International Society of the Learning Sciences.

Latour, B. (2005). *Reassembling the social: An introduction to actor-network-theory*. New York: Oxford University Press.

Lemke, J. L. (2000). Across the scales of time: Artifacts, activities, and meanings in ecosocial systems. *Mind, Culture & Activity*, *7*(4), 273–290.

Licoppe, C., & Smoreda, Z. (2005). Are social networks technologically embedded? How networks are changing today with changes in communication technology. *Social Networks*, *27*(4), 317–335.

Martínez, A., Dimitriadis, Y., Gómez-Sánchez, E., Rubia-Avi, B., Jorrín-Abellán, I., & Marcos, J. A. (2006). Studying participation networks in collaboration using mixed methods. *International Journal of Computer-Supported Collaborative Learning*, *1*(3), 383–408.

Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. Oxford, UK: Oxford University Press.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford, UK: Oxford University Press.

O'Reilly, T. (2005). What is Web 2.0: Design patterns and business models for the next generation of software. http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html

Renninger, K. A., & Shumar, W. (2002). *Building virtual communities: Learning and change in cyberspace*. Cambridge, UK: Cambridge University Press.

Rosé, C. P., Wang, Y.-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *International Journal of Computer-Supported Collaborative Learning*, 3(3), 237–271. doi:10.1007/s11412-007-9034-0

Rosen, D., & Corbit, M. (2009). Social network analysis in virtual environments. *Proceedings of the 20th ACM conference on Hypertext and Hypermedia* (HT '09), 29 June–1 July 2009, Torino, Italy (pp. 317–322). New York: ACM.

Siemens, G. (2013). Massive open online courses: Innovation in education? In R. McGreal, W. Kinuthia, S. Marshall, & T. McNamara (Eds.), *Open educational resources: Innovation, research and practice* (pp. 5–15). Vancouver, BC: Commonwealth of Learning and Athabasca University.

Suthers, D. D. (2006). Technology affordances for intersubjective meaning-making: A research agenda for CSCL. *International Journal of Computer-Supported Collaborative Learning*, 1(3), 315–337.

Suthers, D. D. (2015). From micro-contingencies to network-level phenomena: Multilevel analysis of activity and actors in heterogeneous networked learning environments. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK 'LA), 16–20 March, Poughkeepsie, NY, USA (pp. 368–377). New York: ACM.

Suthers, D. D. (2017). Applications of cohesive subgraph detection algorithms to analyzing socio-technical networks. *Proceedings of the 50th Hawaii International Conference on System Sciences* (HICSS-50), 4–7 January 2017, Waikoloa Village, HI, USA (CD-ROM). IEEE Computer Society.

Suthers, D. D., & Dwyer, N. (2015). Identifying uptake, sessions, and key actors in a socio-technical network. *Proceedings of the 44th Hawaii International Conference on System Sciences* (HICSS-44), 5–8 January 2011, Kauai, Hawai'i (CD-ROM). IEEE Computer Society.

Suthers, D. D., Dwyer, N., Medina, R., & Vatrapu, R. (2010). A framework for conceptualizing, representing, and analyzing distributed interaction. *International Journal of Computer Supported Collaborative Learning*, 5(1), 5–42. doi:10.1007/s11412-009-9081-9

Suthers, D. D., Fusco, J., Schank, P., Chu, K.-H., & Schlager, M. (2013). Discovery of community structures in a heterogeneous professional online network. *Proceedings of the 46th Hawaii International Conference on the System Sciences* (HICSS-46), 7–10 January 2013, Maui, HI, USA (CD-ROM). IEEE Computer Society.

Suthers, D. D., Lund, K., Rosé, C. P., Teplovs, C., & Law, N. (2013). *Productive multivocality in the analysis of group interactions*. Springer.

Suthers, D. D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning. In G. Conole, D. Gašević, P. Long, & G. Siemens (Eds.), *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK 'LA), 27 February–1 March 2011, Banff, AB, Canada (pp. 64–74). New York: ACM.

Tönnies, F. (2001). *Community and civil society* (J. Harris & M. Hollis, Trans. from Gemeinschaft und Gesellschaft, 1887) Cambridge, UK: Cambridge University Press.

Trausan-Matu, S., & Rebedea, T. (2010). A polyphonic model and system for inter-animation analysis in chat conversations with multiple participants. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (pp. 354–363). Springer.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. New York: Cambridge University Press.

Wellman, B., Quan-Haase, A., Boase, J., Chen, W., Hampton, K., Diaz, I., & Miyata, K. (2003). The social affordances of the Internet for networked individualism. *Journal of Computer-Mediated Communication*, 8(3). doi:10.1111/j.1083-6101.2003.tb00216.x

Chapter 17: Data Mining Large-Scale Formative Writing

Peter W. Foltz[1,2], Mark Rosenstein[2]

[1] Institute of Cognitive Science, University of Colorado, USA
[2]Advanced Computing and Data Science Laboratory, Pearson, USA
DOI: 10.18608/hla17.017

ABSTRACT

Student writing in digital educational environments can provide a wealth of information about the processes involved in learning to write as well as evidence for the impact of the digital environment on those processes. Developing writing skills is highly dependent on students having opportunities to practice, most particularly when they are supported with frequent feedback and are taught strategies for planning, revising, and editing their compositions. Formative systems incorporating automated writing scoring provide the opportunities for students to write, receive feedback, and then revise essays in a timely iterative cycle. This chapter provides an analysis of a large-scale formative writing system using over a million student essays written in response to several hundred pre-defined prompts used to improve educational outcomes, better understand the role of feedback in writing, drive improvements in formative technology, and design better kinds of feedback and scaffolding to support students in the writing process.

Keywords: Writing, formative feedback, automated scoring, mixed effects modelling, visualization, writing analytics

Writing is an integral part of educational practice, where it serves both as a means to train students how to express knowledge and skills as well as to help improve their knowledge. It is well established that in order to become a good writer, students need a lot of practice. However, just practicing writing is insufficient to become a good writer; receiving timely feedback is critical (e.g., Black & William, 1998; Hattie & Timperley, 2007; Shute, 2008). Studies of formative writing in the classroom (e.g., Graham, Harris, & Hebert, 2011; Graham & Hebert, 2010; Graham & Perin, 2007) have shown that supporting students with feedback and providing instruction in strategies for planning, revising, and editing their compositions can have strong effects on improving student writing.

Text as Data

Writing is a complex activity and can be considered a form of performance-based learning and assessment, in that students are performing a task similar to what they will typically be expected to carry out in their future academic and work life. As such, writing provides a rich source of data about student content knowledge, expressive skills, and language ability. Thus, writing affords making multiple inferences about the nature of student performance based on the textual information.

Currently most writing is mediated by computer, which provides an opportunity to study and impact writing learning at a depth and over time periods that were just not practical with paper-based media. For instance, Walvoord and McCarthy (1990), with a series of collaborators, conducted classroom studies over nearly a decade, gathering artifacts such as student journals, drafts, and final papers to build understandings of writing instruction. Much of the effort to conduct the study was in the collection and hand analyses. Today, with computer-based writing, such resources are more readily available as part of the writing process, and are in a form where natural language processing and machine learning can be automatically employed. By applying appropriate learning analytic methods, textual information can therefore be automatically converted to data to support inferences about student performance.

Automated analyses have been applied to understanding aspects of writing since the 1960s. Content analysis (e.g., Gerbner, Holsti, Krippendorff, Paisley, & Stone, 1969; Krippendorff & Bock, 2009) was designed to allow analysis of textual data in order to make replicable, valid inferences about the content. However, the methods focused primarily on counts of key terms used in the texts. Ellis Page (1967) pioneered techniques to convert the language features of student writing into scores that correlated highly with teacher ratings of the essays. With the advent of increasingly more sophisticated natural language processing and machine learning techniques over the past 50 years, automated essay scoring (AES) has now become a widely used set of approaches that can provide scores and feedback instantly. Research on AES systems has shown that their scoring can be as accurate as human scorers (e.g., Burstein, Chodorow, & Leacock, 2004; Landauer, Laham, & Foltz, 2001; Shermis & Hamner, 2012), can score multiple traits of writing (e.g., Foltz, Streeter, Lochbaum, & Landauer, 2013), and can be used for feedback on content (e.g., Foltz, Gilliam, & Kendall, 2000; Foltz et al., 2013).

While much of the focus in the evaluation of AES has examined the accuracy of the scoring and the different types of essays that can be scored, AES also has wide applicability to formative writing, where evaluation can focus more on how it aids student learning. Human assessment of writing can be time consuming and subjective, limiting the opportunities for students to receive feedback. As a component of a formative tool, AES can provide instantaneous feedback to students and support the teaching of writing strategies based on detecting the types of difficulties students encounter. For example, when incorporated into classroom instruction, students are able to write, submit, receive feedback, and revise essays multiple times over a class period. All student writing is performed electronically, and is automatically scored and recorded, providing a record of all student actions and all feedback they received. This archive permits continuous monitoring of performance changes in individuals as well as across larger groups of students, such as classes or schools. Teachers can analyze the progress of each student in a class and intervene when needed. In addition, it now becomes possible to chart progress across the class in order to measure effectiveness of curricula and teaching strategies as reflected in student writing performance. A number of formative writing tools using automated scoring have been developed and are in use, including WriteToLearn™ (W2L; Landauer, Lochbaum, & Dooley, 2009), Criterion, (Burstein, Chodorow, & Leacock, 2004), OpenEssayist (Whitelock, Field, Pulman, Richardson, & Van Labeke, 2013), and Writing Pal (Roscoe & McNamara, 2013).

Data Mining Applied to Writing

Automated formative assessment of writing provides a rich data set to examine the changes in writing performance and system features that influence that performance. With the increasing adoption of digital educational environments, there are new opportunities to leverage the data from student interactions in these environments as evidence (e.g., DiCerbo & Behrens, 2012). Recent work has begun to extract and analyze writing from writing assignments, from peer grading exercises, as well as from collaborative forum discussions in order to examine student performance. While there have been a number of overviews of data mining methods (e.g., Peña-Ayala, 2014; Romero & Ventura, 2007; Romero & Ventura, 2013), there has still been little focus on large-scale data mining of formative writing. With the advent of more powerful computational discourse tools, new techniques are emerging (e.g., Buckingham-Shum, 2013; McNamara, Allen, Crossley, Dascalu, & Perret, this volume; Rosé, this volume).

Some studies have examined large corpora of student writing, although not focused on the aspects of formative feedback. For example, Parr (2010) analyzed 20,000 essays written to 60 different prompts at different grade levels in order to measure how writing skills develop for different genres of essays. All scoring of the essays was performed by human scorers, although tools were provided to make the scoring easier and to ensure consistency. Deane and Quinlan (2010) performed analyses using the e-Rater automated scoring engine to extract features from thousands of essays and then performed factor analysis in order to examine developmental levels and linguistic dimensions of writing. Deane (2014) also used automated scoring of essays from a multi-state implementation, analyzing features from keystroke logs and the essays themselves, in order to predict factors of writing ability and reading level.

Aspects of the formative process have also been examined using smaller samples of data; for example, the research on collaborative writing at the University of Sidney (Calvo, O'Rourke, Jones, Yacef, & Reimann, 2011; Reimann, Calvo, Yacef, & Southavilay, 2010) used student log data and automated assessment to support writing. In their work, they analyzed grammatical and topical aspects of writing as well as log files of the sequences of revisions and writing activities in order to understand team writing processes. In addition, research has performed fine-grained analysis of writing by coupling log data with physiological monitoring, such as eye tracking (e.g., Leijten & Van Waes, 2013). WhiteLock et al. (2013, 2015) used visualizations of textual features of essays, including displays of key words and phrases and information about essay structure across multiple essays as a way to allow students and instructors to understand aspects of the content of the essays. These visualizations can then be used as the basis for providing advice for improving student writing.

Other research involving writing and data mining has considered writing as a secondary task, such as Crossley et al. (2015), who examined student writing in discussion forums within MOOCs to predict whether a student would successfully complete the course, and White and Larusson (2014), who developed lexical analysis techniques to analyze changes in student writing to detect when students reach the point when they sufficiently understand a core concept in order to re-express it in their own words. Finally, analyses of feedback during the revision process in online systems (e.g., Baikadi, Schunn, & Ashley, 2015; Calvo, Aditomo, Southavilary, & Yacef, 2012) has shown what kinds of feedback can be most effective in the revision process. The majority of these studies focused on analyses based on tens to hundreds of students, so while they inform the use of data mining techniques and provide critical information on the role of formative feedback, they have not yet been scaled larger administrations.

This chapter builds on the above approaches to describe an approach to large-scale analysis of writing by applying data mining to components of the formative writing process on hundreds of thousands to over a million samples of writing collected from a formative online writing system. The analyses are used to investigate specific classes of questions about how a formative system is currently being used, its efficacy, and how understanding current use yields suggestions for improved learning, both through improving the system implementation and by introducing direct interventions aimed at students using the system. The chapter illustrates approaches utilizing descriptive statistics of performance as well as formally modelling changes in performance. While the chapter focuses on methodology, the intent is to illustrate how writing data can be used more generally to inform decisions about the quality of student learning, about the effectiveness of implementation in the classroom, as well as the effectiveness of the digital environment itself as an educational tool.

ONLINE FORMATIVE WRITING SYSTEM

The context used to illustrate the power of data mining in the lifecycle of a large-scale implementation was conducted with student interaction data from the formative writing assessment system WriteToLearn™. WriteToLearn™ is a web-based writing environment that provides students with exercises to write responses to narrative, expository, descriptive, and persuasive prompts as well as to read and write summaries of texts in order to build reading comprehension. Students use the software as an iterative writing tool in which they write, receive feedback, and then revise and resubmit their improved essays. The automated feedback provides an overall score and individual trait scores such as "ideas, organization, conventions, word

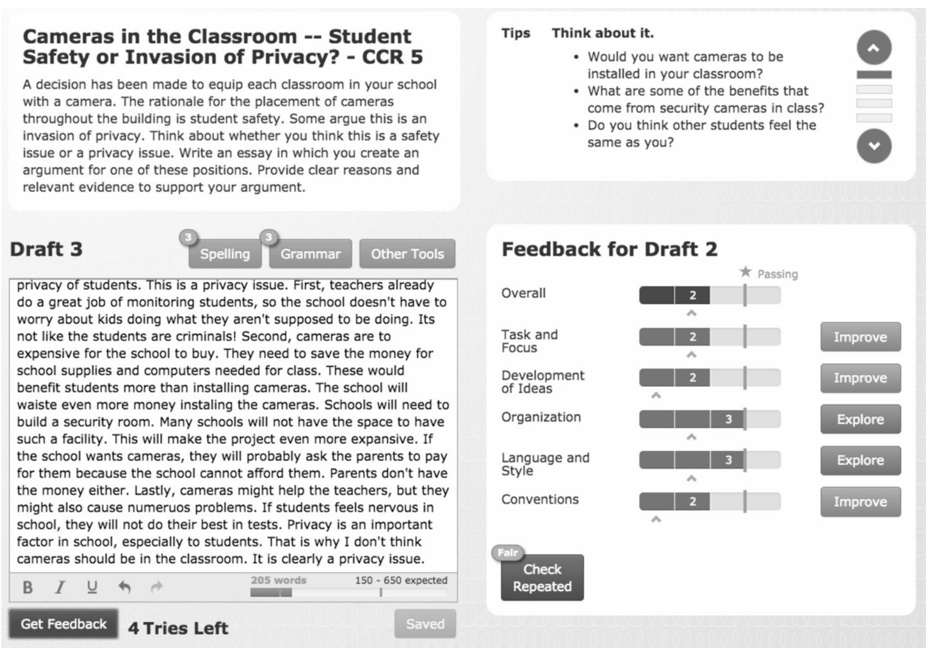


Figure 17.1. Essay Feedback Scoreboard. WriteToLearn™ feedback with an overall score, scores on six popular traits of writing, as well as support for the writing process.

choice, and sentence fluency." The student can also view supplemental educational material to help them understand the feedback, as well as suggest approaches to improve their writing. In addition, grammar and spelling errors are flagged. Figure 1 shows a portion of the system's interface, in this case illustrating the scoring feedback resulting from a submission to a 12th grade persuasive prompt. Evaluations of WriteToLearn™ have shown significantly better reading comprehension and writing skills resulting from two weeks of use (Landauer et al., 2009) as well as validating the system scores being as reliable as human raters, and significantly improved end-of-year pass rates on a statewide writing assessment (Mollette & Harmon, 2015).

Algorithm for Scoring Writing

WriteToLearn's™ automated scoring is based on an implementation of the Intelligent Essay Assessor (IEA). IEA is trained to associate extracted features from each essay to scores assigned by human scorers. A machine-learning-based approach is used to determine the optimal set of features and the weights for each of the features to best model the scores for each essay. From these comparisons, a prompt- and trait-specific scoring model is derived to predict the scores that the same scorers would assign to new responses. Based on this scoring model, new essays can be immediately scored by analysis of the features weighted according to the scoring model. The focus in this chapter is not on the actual algorithms or features that make up the scoring, as those have been described in detail elsewhere (see Landauer et al., 2001; Foltz et al., 2013). Instead, the focus is how the trail left by automated scoring and student actions can be used to monitor learning across large sets of writing data and facilitate improvements in the formative system.

Data

The data comprised two large samples of student interactions with WriteToLearn™ collected from U.S. adoptions of the software. One set comprised approximately 1.3 million essays from 360,000 assignments written by 94,000 students collected over a 4-year period. The second set represented approximately 62,000 student sessions with nearly 900,000 actions. The data included student essays and a time-stamped log of all student actions, revisions, and feedback given by the system. Essays were recorded each time a student submitted or saved an essay, resulting in a record of each draft submitted. The essays were written to approximately 200 pre-defined prompts. No human scoring was performed on these essays. All essay scores were generated by automated scoring, with the prediction performance of the models validated against human agreement from test sets or using cross-validation.

Analyses Enabled by the Approach

At all stages in the lifecycle of a formative system — design, implementation, deployment, redesign, and maintenance — analysis of actual use via analytics applied to log data can inform improvements to the system. As Mislevy, Behrens, Dicerbo, and Levy (2012) note, there is interplay between evidence-centred design, which represents best practices when a system is first conceived and data mining student actions of the implementation that reflects actual use, where each is critical in building and evolving educational systems. From the design phase, we are interested in analyzing use data to assess our assumptions and, in our case, determining if cycles of writing, feedback, and revising improves writing performance and at what rate and whether the rate of improvement differs among the traits of writing. In terms of pedagogical theory, we want to understand what mix of writing, mechanics feedback, content feedback, and revising leads to optimal learning, and potentially how to individualize advice to students and teachers. Currently the system by default allows six revision/feedback cycles with teachers able to customize the limit, and use data should help develop guidelines for this feature. Another quite productive form of analysis is to model student performance; here we discuss a mixed effects model that allows us to estimate the relative difficulty of the prompts. Prompts typically are assigned a grade level when developed, but modelling allows us to determine if the prompt is correctly labelled; using performance data from millions of essays written to a prompt allows finer grained levelling.

Many additional types of analysis are possible with writing-log data than there is room to detail in this chapter (see also Calvo et al., 2012; Deane, 2014). Two areas we have found particularly promising are evaluation of teachers' instructional strategies; for instance, in terms of which prompts were chosen and how long (a single class period, a week, longer) students were allowed to write to a prompt. While systems such as the one described here have professional development instruction for teachers as well as teachers' guides, it is astonishingly useful to observe how the system is actually used in classrooms in order to uncover new strategies and measure the relative effectiveness among the strategies. Another area that we lack space to describe in detail is fine-grained analysis of student actions. For instance, it is possible to tell when and where in the writing process a student exploits a help facility, and often possible to infer when a student should have taken advantage of a facility but didn't — from which it may be possible to infer redesign choices in terms of user interface layout and other design issues. Additional discussion of some of these directions can be found in Foltz and Rosenstein

(2013), Foltz and Rosenstein (2015), and illustrated in Foltz and Rosenstein (2016).

VALIDATING THEORY

Does Writing and Revising Result in Improved Writing Performance?

Formative writing systems are designed to support a rapid cycle of *write*, *submit*, *receive feedback*, and *revise*. This cycle is one of the key differentiators of automated formative writing from standard classroom writing practice, where human scoring of essays is time consuming so students cannot receive immediate feedback. Thus, it is critical to determine how often students submit and revise essays and determine the factors and time paths that lead to greatest success. This can help address questions of whether revising results in better writing, as measured by the automated scores and what patterns of use facilitate the most rapid improvement.

Using a subset of the data, we examined writing over a single semester in which teachers in three grades (5th, 7th, and 10th) across an entire state assigned writing exercises to students. During that period, 21,137 students wrote to 72,051 assignments (an average of almost four assignments per student) with 107 different unique writing prompts assigned. These assignments resulted in 255,741 essays submitted and scored over the period of analysis. For each submission, students received feedback and scores on their overall essay quality, as well on six different writing traits: ideas, organization, conventions, word choice, sentence fluency, and voice. While there was a wide distribution in the number of revisions students made, most students made more than one revision, with most making up to five revisions. Figure 2 shows the score improvement (score on last attempt minus score on first attempt) for students who wrote multiple drafts. It shows improvement for each of the six writing traits as well as the overall score. There is a clear trend indicating that more revisions equal higher scores. With the typical five revisions, the average student score improved by almost one score point (out of a maximum of 6). Generally, we see greatest improvement in scores for content-based features, such as ideas, voice, and organization, and less for features related to writing skill, such sentence fluency and writing conventions. The smoothness of the curves and small error bars are due to the large number of data points for each revision from 0 to 5.

Time Spent Between Revisions

We can further investigate the impact on student performance of the time-spent writing before requesting feedback to better understand the best allocation of time among the *write*, *submit*, *feedback*, and *revise* phases. Using data from approximately 1.1 million student writing attempts across a wide range of users of WriteToLearn™, we calculated the change in student grade (e.g., improvement from one draft to the next) based on how much time was spent between drafts. The change in grade shown in Figure 3 indicates that the improvement in writing score generally increases up to about 25 minutes at which point it levels off and begins to drop. In addition, most of the nega-

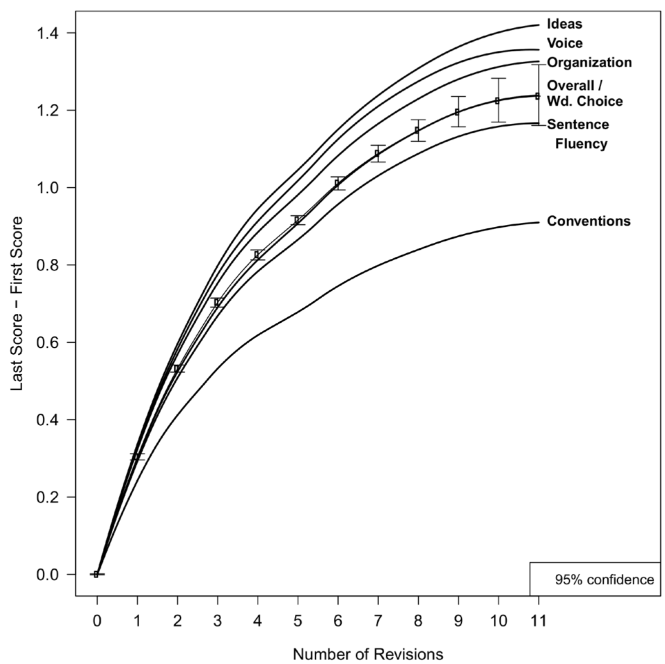


Figure 17.2. Change in writing scores for multiple writing traits across revisions.

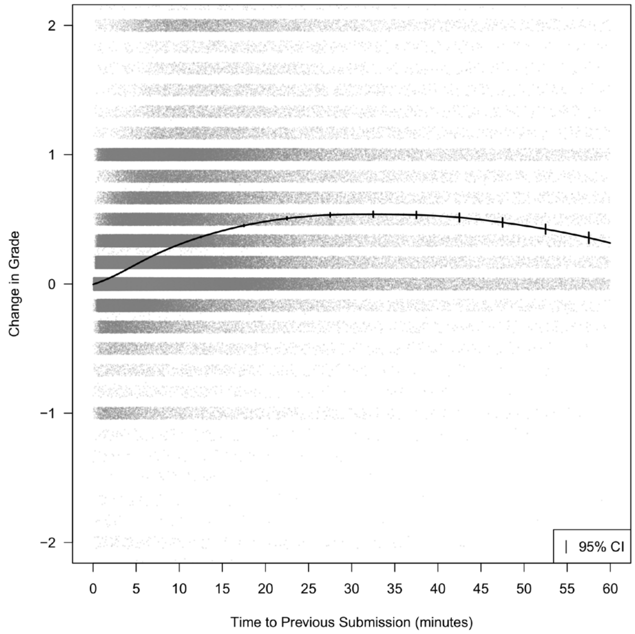


Figure 17.3. Change in grade over revisions based on the time to revise.

tive change (essays receiving a lower score than the previous version) occurs with revisions of less than five minutes. The results suggest an optimal range of time to spend revising before requesting additional feedback. These two results indicate how analysis of log data can validate that the write-feedback-revise cycle improves writing skills, as well as illustrates the ability to fine-tune learning by attempting to lead the student into more effective cycles where feedback is requested at appropriate intervals.

Modelling

The underlying structure of the writing process, as it manifests within a formative writing tool, is often best made interpretable through the construction of formal statistical models. With their explicit representations of the complex interplay of revising, receiving writing advice, and composing responses to multiple prompts over time, these models provide estimates and confidence intervals for parameters of critical interest. Grounded by the student log data, these models can account and control for the complex covariance structure implicit within this stream of data with its aspects of repeated measures of performance received on shared prompts embedded in an overall longitudinal model of growth that can span a significant portion of the total time a student receives writing instruction. A carefully constructed model facilitates teasing out student progress with exposure to the tool, allows placing both students and items on scales of skill level and difficulty respectively, and provides estimates on how changing levels of exposure to components of the available feedback impacts writing performance.

The models described here are based on over 840,000 essays written against more than 190 prompts over a 4-year period by approximately 80,000 students, where over 20% of the students were followed for three or more years. The models predict the holistic score for each essay submitted for feedback, which given the explanatory variables signifies the expected score a student would receive on their essay. The explanatory variables allow us to estimate and control for factors such as the student’s grade level, the length of the essay, and the difficulty of the prompt.

The writing process is represented within a linear mixed effects model framework (Pinheiro & Bates, 2006), building on the techniques described in Baayen, Davidson, and Bates (2008). Mixed effects models can estimate both the student’s "skill level" and the item's “difficulty” by viewing them as being sampled from a population of all potential students and a bank of all potential prompts, estimates computed in addition to the relationships that hold over the entire population. The students and prompts were modelled as random effects drawn from a distribution with a mean of zero and with the standard deviation estimated from the data. The derived variability provides an estimate of student individual differences, while also capturing the variability of item difficulty. Table 1 contains descriptions of the fixed and random effects used in the models.

At each student grade level, the impact of the higher grade is to increase the score, while as content grade level increases (the labelled grade level of a prompt) the expected score decreases. Finally, in controlling

Table 17.1. Description of Fixed and Random Effects

Variable Name	Description
Fixed Effects	
studentGradeLevel:n	Student's grade level as a factor level (coefficient is the difference between grade n and grade 3)
contentGradeLevel	Grade level of prompt (an assigned level)
log10(wordCount)	Log base 10 of word length of essay
attempt	For a given prompt, the revision of this specific essay submission
elapsedTimeDay	A measure of time in days of how long since first W2L use (a measure of age-based growth)
cumW2LTimeDay	Total face-time student has had with W2L by this submission
interaction	Total number of submissions this student has made to W2L
Random Effects	
studentID	Factor levels, one for each student
contentID	Factor levels, one for each prompt

for essay length, a longer essay on average would be expected to receive a higher score. The four measures of exposure to WriteToLearn™ are all statistically significant and positive, indicating its cumulative positive effect.

While the four measures of WriteToLearn™ are related – e.g., as number of attempts on a specific prompt increases, concurrently the total time spent using WriteToLearn™ increases – they capture different aspects of student interaction with the system. The effect sizes seems small; for instance, each additional attempt on a specific prompt increases the expected score by only .018, a number that represents just the increase based on receiving feedback on a single revision of the essay. In fact, it is only through data mining and modelling with large data sets that we can reliably estimate these important small, incremental effects. From a more global perspective, the cumulative impact of attempts and time spent interacting with WriteToLearn™ result in improvements in achievement. This progress is often best benchmarked with external validations such as those observed in improved pass rates on state achievement tests with more intensive use (Mollette & Harmon, 2015).

Modelling to Determine Writing Prompt Difficulty

Many pedagogical considerations arise in assigning a prompt to a student or a class and one often-expressed concern is adjusting the scoring of the prompt to the student's level (see also Deane & Quinlan, 2010, for a related approach to determining prompt difficulty from a writing corpus). Although some prompts require a threshold skill level or specific knowledge or expertise to be addressed, many are applicable for students over a wide grade range. What differs in the assignment is the expectation of the quality or skill evidenced by the final product and its evaluation via a score. Scoring of prompts is based on grade-specific models, so a prompt labelled as appropriate for 10th graders implies that it is both well-suited for the skills and knowledge expected in 10th grade, but also that the automated scoring was calibrated using training-set essays written by 10th graders. In cases where a prompt is appropriate for a range of grade levels, and training sets of students at different grade levels were available, the same prompt may appear at multiple grade levels, where the critical difference is that different scoring models are used to evaluate the student's work at each grade level.

Often teachers prefer finer levels of discrimination among prompts, such as having a measure of the relative difficulty of a set of prompts that fit the grade level of their class. This is exactly the case that the random effect estimates of the prompts can be used to address. As a prompt's labelled grade level increases, the coefficient on fixed effect contentGradeLevel in the model indicates a 0.073 decrease in expected score (harder prompts contribute to lower scores), other variables held constant. Equivalently, controlling for the labelled prompt grade level, the individual prompt random effects indicates how strongly a given prompt differs in difficulty from this mean fixed effect. This allows ordering the prompts within grade levels, providing empirically derived additional support infrastructure for teachers. Similarly, taking into account the fixed prompt effect allows ordering all of the prompts, which broadens the set of prompts a teacher may be comfortable assigning.

Title	Grade Level	Difficulty
Freedom of Speech	12	0.80
How to Start a Hobby – A/B	6	0.76
Essay About Causes and Effects in History	11	0.72
How to Start a Hobby	5	0.71
How to Start a Hobby	6	0.70
American President	10	0.68
What's Cooking?	6	0.66
Consumer Reporter	12	0.64
What's Cooking? – A/B	6	0.63
Favorite Activity	4	0.63
...		
Effects of Texting on Communication Skills	8	−0.70
Should Recycling be Voluntary or Required?	7	−0.70
How Much Time to Play Computer Games	7	−0.71
An Unusual Event	9	−0.74
An Important Decision	8	−0.75
Interpret a Literary Theme	7	−0.79
A Meaningful Childhood Memory	10	−0.81
A Meaningful Life Lesson	10	−0.82
Dealing with Conflict	10	−0.85
Compare and Contrast Two Literary Characters	10	−1.08

Beyond this practical result, estimates of prompt difficulty controlling for assigned grade level raise a number of interesting research questions. Table 2 presents a subset of the prompts ordered by the estimates of the conditional modes of the random effects (Bates, Maechler, Bolker, & Walker, 2015) shown in the column called difficulty along with columns for the labelled grade level and the prompt title. The impact on score received on an essay is the sum of the grade level times its coefficient from the model plus its difficulty, so the more positive difficulty is, the easier the prompt is relative to other prompts at that grade level; hence, prompts near the bottom of the table, relative to their grade level, are more difficult. We are just in the early stages of trying to form hypotheses to explain this data, such as why the first 10 prompts in the table are so much easier than other prompts at that grade level and why the last 10 are so much harder, as well as why the relatively easiest items seem to be pulled over a broader range of grade levels than the more constrained set of relatively most difficult items.

Considerations in Modelling with Large Data Sets

In designing a model, there are trade-offs between expressiveness and parsimony. With large datasets, often statistical significance is not a sufficient basis to decide on model form; the purpose of the analysis must also be factored into the decision. A strong message from the descriptive plots presented earlier was that of diminishing returns for variables such as number of submissions per essay. This tendency could be described with a polynomial or in a general additive model context. The power of data mining a large data set is that we can make fewer assumptions about the form relationships will take. In this case, we could assume a linear relation between performance and grade, but instead we estimated a separate improvement relative to 3rd grade, as a baseline, and plotted the relationship in Figure 4. Additional research is necessary to better understand the causes of the asymptotic behaviour and the implications for potential improvements to WriteToLearn™.

We see that from the 4^{th} through about the 10^{th} grade, the improvement is approximately linear, but asymptotes out for 11^{th} and 12^{th} grade. This indicates that at least with this set of prompts and their scoring models, WriteToLearn™ has difficulty distinguishing improvement in writing among 10th through 12th graders. Estimating the slope of the linear portion of the curve from grades 4 to 10 yields a gain of 0.048/grade level, which also can be expressed as an expected gain of 0.29 in going from 4^{th} to 10^{th} grade. This is the expected gain, holding the use of WriteToLearn™ constant. Additional research is necessary to better understand the causes of the asymptotic behaviour and potential improvements to WriteToLearn™.

Related to the work described here are finer grained models of action transitions also using mixed effect models (for instance in a tutoring context see Feng, Heffernan, Heffernan, & Mani, 2009) or using Markov methods (e.g., Beal, Mitra, & Cohen, 2007; Jeong et al., 2008) or Bayesian techniques (e.g., Conati et al., 1997). These techniques can be used to better understand student interactions at the action level (such as use of scaffolding facilities) that complement the more course grained analysis described here.

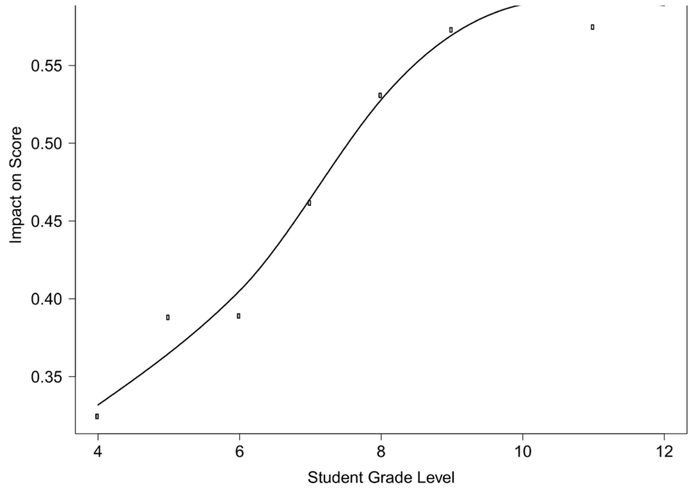


Figure 17.4. Improvement in student score by grade level.

CONCLUSION

Digital education environments can provide an infrastructure to support students with more personalized learning experiences by having students work on more authentic educational tasks while receiving immediate feedback and training specific to their learning needs. Properly instrumented, these environments can also provide a rich source of information about student learning and progress as they interact with the system. Large-scale implementations of formative writing provide rich sets of data for analysis of performance and effects of feedback. By treating the written product as data, applying automated scoring of writing allows monitoring of student learning as students write and revise essays within these implementations. By examining the log of student actions, the amount of time taken, and the changes in the essays, one can track the impact on learning from use of the system.

Developing and maintaining a formative system in a manner to maximize student learning growth requires a range of decisions be made starting from the design and implementation and continuing through the monitoring of its use. Decisions in the design and implementation phase are typically limited to theory and best practices, which are often at a level of granularity that affords a great deal of ambiguity in implementation. However, once a system is deployed, these assumptions can be cast against the actual behaviour of teachers applying the system during their classroom activities and students learning to write. Through data mining, these assumptions can be tested, both to validate the assumptions of the system and to gain greater insight into how students learn.

Writing to Learn and Learning to Write

The resulting analysis validates a key tenet of formative writing: students can improve their writing with revisions based on feedback from the system. A data mining approach to writing permits a fine-grained approach to examining the changes in learning and the effects of feedback on performance. This further permits us to discover, prioritize, and address concerns as they arise and determine which changes are most likely to improve the student experience and their ability to sharpen their writing skills.

The focus of writing assessment has often been put on the product (i.e., the final essay). By performing data mining on student draft submissions and the log of their actions, it is possible to track the process that learners take to create the product. This analysis allows interventions to be performed at strategic points during the process of writing rather than just evaluating the end-product. A wide range of types of analyses can be performed on writing data, including examining the essays, the process to create the essays, as well as the progress of the changes. These approaches can be both descriptive analyses and modelling. While we could not possibly provide a comprehensive discussion on all types of analyses in this chapter, the goal was to illustrate a variety of approaches to show how data mining can provide new ways of thinking about collecting evidence of system and student writing performance and uncover patterns that go beyond those apparent from only observing individual students or classrooms.

REFERENCES

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390–412.

Baikadi, A., Schunn, C., & Ashley, K. (2015). Understanding revision planning in peer-reviewed writing. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Education Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 544 – 548). International Educational Data Mining Society.

Bates, D., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. ArXiv e-print, *Journal of Statistical Software*, http://arxiv.org/abs/1406.5823.

Beal, C., Mitra, S., & Cohen, P. R. (2007). Modeling learning patterns of students with a tutoring system using Hidden Markov Models. *Proceedings of the 2007 conference on Artificial Intelligence in Education: Building Technology Rich Learning Contexts That Work*, 238–245.

Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7–74.

Buckingham-Shum, S. (2013). *Proceedings of the 1st International Workshop on Discourse-Centric Analytics*, workshop held in conjunction with the 3rd International Conference on Learning Analytics and Knowledge (LAK '13), 8–12 April 2013, Leuven, Belgium. New York: ACM.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The Criterion Online writing service. *AI Magazine*, 25(3), 27–36.

Calvo, R. A., Aditomo, A., Southavilay, V., & Yacef, K. (2012). The use of text and process mining techniques to study the impact of feedback on students' writing processes. *Proceedings of the 10th International Conference of the Learning Sciences* (ICLS '12) Vol. 1, Full Papers, 2–6 July 2012, Sydney, Australia (pp. 416–423).

Calvo, R. A., O'Rourke, S. T., Jones, J., Yacef, K., & Reimann, P. (2011). Collaborative writing support tools on the cloud. *IEEE Transactions on Learning Technologies*, 4(1), 88–97.

Conati, C., Gertner, A. S., VanLehn, K., & Druzdzel, M. J. (1997). On-line student modeling for coached problem solving using Bayesian networks. *Proceedings of the 6th International User Modeling Conference* (UM97) (pp. 231–242).

Crossley, S. A., McNamara, D. S., Baker, R., Wang, Y., Paquette, L., Barnes, T., & Bergner, Y. (2015). Language to completion: Success in an educational data mining massive open online class. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Education Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 388–392). International Educational Data Mining Society.

Deane, P. (2014). Using writing process and product features to assess writing quality and explore how those features relate to other literacy tasks. *Educational Testing Research Report* ETS RR-14-03. http://dx.doi.org/10.1002/ets2.12002.

Deane, P., & Quinlan, T. (2010). What automated analyses of corpora can tell us about students' writing skills. *Journal of Writing Research*, 2(2), 151–177.

DiCerbo, K. E., & Behrens, J. (2012). Implications of the digital ocean on current and future assessment. In R. Lissitz & H. Jao (Eds.), *Computers and their impact on state assessment: Recent history and predictions for the future*. Charlotte, NC: Information Age.

Feng, M., Heffernan, N. T., Heffernan, C., & Mani, M. (2009). Using mixed-effects modeling to analyze different grain-sized skill models in an intelligent tutoring system. *IEEE Transactions on Learning Technologies*, 2, 79–92.

Foltz, P. W., Gilliam, S., & Kendall, S. (2000). Supporting content-based feedback in online writing evaluation with LSA. *Interactive Learning Environments*, 8(2), 111–129.

Foltz, P. W., & Rosenstein, M. (2013). Tracking student learning in a state-wide implementation of automated writing scoring. *Proceedings of the Neural Information Processing Systems* (NIPS) Workshop on Data Driven

Education. http://lytics.stanford.edu/datadriveneducation/

Foltz, P. W., & Rosenstein, M. (2015). Analysis of a large-scale formative writing assessment system with automated feedback. *Proceedings of the 2nd ACM conference on Learning@Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 339–342). New York: ACM.

Foltz, P. W., & Rosenstein, M. (2016). Visualizing teacher assignment behavior in a statewide implementation of a formative writing system. Cover competition. *Education Measurement: Issues and Practice*, *35*(2), 31. http://dx.doi.org/10.1111/emip.12114

Foltz, P. W., Streeter, L. A., Lochbaum, K. E., & Landauer, T. K. (2013). Implementation and applications of the Intelligent Essay Assessor. In M. D. Shermis & J. Burstein, *handbook of automated essay evaluation: Current applications and future directions* (pp. 68–88). New York: Routledge.

Gerbner, G., Holsti, O. R., Krippendorff, K., Paisley, W. J., & Stone, Ph. J. (Eds.) (1969). The analysis of communication content: Development in scientific theories and computer techniques. New York: Wiley.

Graham, S., Harris, K. R., & Hebert, M. (2011). *Informing writing: The benefits of formative assessment*. Carnegie Corporation of New York.

Graham, S., & Hebert, M. (2010). *Writing to read: Evidence for how writing can improve reading*. Carnegie Corporation of New York.

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445–476.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.

Jeong, H., Gupta, A., Roscoe, R., Wagster, J., Biswas, G., & Schwartz, D. (2008). Using Hidden Markov Models to characterize student behaviors in learning-by-teaching environments. In B. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *Proceedings of the 9th International Conference on Intelligent Tutoring Systems* (ITS 2008), 23–27 June 2008, Montreal, PQ, Canada (pp. 614–625). Berlin/Heidelberg: Springer.

Krippendorff, K., & Bock, M. A. (2009). *The content analysis reader*. Sage Publications.

Landauer, T. K., Laham, D., & Foltz, P. W. (2001). Automated essay scoring. *IEEE Intelligent Systems*, September/October.

Landauer, T., Lochbaum, K., & Dooley, S. (2009). A new formative assessment technology for reading and writing. *Theory into Practice*, 48(1), 44–52. http://dx.doi.org/10.1080/00405840802577593

Leijten, M., & Van Waes, L. (2013). Keystroke logging in writing research using Inputlog to analyze and visualize writing processes. *Written Communication*, *30*(3), 358–392.

Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining*, 4(1), 11–48.

Mollette, M., & Harmon, J. (2015). Student-level analysis of Write to Learn effects on state writing test scores. Paper presented at the 2015 annual meeting of the American Educational Research Association. http://www.aera.net/Publications/Online-Paper-Repository/AERA-Online-Paper-Repository

Page, E. B. (1967). The imminence of grading essays by computer. *Phi Delta Kappan*, 47, 238–243.

Parr, J. (2010). A dual purpose data base for research and diagnostic assessment of student writing. *Journal of Writing Research*, 2(2), 129–150.

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert systems with applications*, 41(4), 1432–1462.

Pinheiro, J., & Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. New York: Springer-Verlag.

Reimann, P., Calvo, R., Yacef, K., & Southavilay, V. (2010). Comprehensive computational support for collaborative learning from writing. In S. L. Wong, S. C. Kong, & F.-Y. Yu (Eds.), *Proceedings of the 18th International Conference on Computers in Education* (ICCE 2010), 29 November–3 December, Putrajaya, Malaysia (pp.

129–136). Asia-Pacific Society for Computers in Education.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27.

Roscoe, R., & McNamara, D. S. (2013). Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology*, 105(4), 1010–1025. http://dx.doi.org/10.1037/a0032340

Shermis, M., & Hamner, B. (2012). Contrasting state-of-the-art automated scoring of essays: Analysis. Paper presented at Annual Meeting of the National Council on Measurement in Education, Vancouver, Canada, April.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189.

Walvoord, B. E., & McCarthy, L. P. (1990). *Thinking and writing in college*: A *naturalistic study of students in four disciplines*. Urbana, IL: National Council of Teachers of English.

White, B., & Larusson, J. A. (Eds.). (2014). *Learning analytics: From research to practice*. New York: Springer Science+Business Media. doi:10.1007/978-1-4614-3305-7_8.

Whitelock, D., Field, D., Pulman, S., Richardson, J. T. E., & Van Labeke, N. (2013). OpenEssayist: an automated feedback system that supports university students as they write summative essays. *Proceedings of the 1st International Conference on Open Learning: Role, Challenges and Aspirations*. The Arab Open University, Kuwait, 25–27 November 2013.

Whitelock, D., Twiner, A., Richardson, J. T. E., Field, D., & Pulman, S. (2015). OpenEssayist: A supply and demand learning analytics tool for drafting academic essays. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March, Poughkeepsie, NY, USA (pp. 208–212). New York: ACM.

Chapter 18: Diverse Big Data and Randomized Field Experiments in MOOCs

René F. Kizilcec,[1] Christopher Brooks[2]

[1]Department of Communication, Stanford University, USA
[2]School of Information, University of Michigan, USA
DOI: 10.18608/hla17.018

ABSTRACT

A new mechanism for delivering educational content at large scale, massive open online courses (MOOCs), has attracted millions of learners worldwide. Following a concise account of the recent history of MOOCs, this chapter focuses on their potential as a research instrument. We identify two critical affordances of this environment for advancing research on learning analytics and the science of learning more broadly. The first affordance is the availability of diverse big data in education. Research with heterogeneous samples of learners can advance a more inclusive science of learning, one that better accounts for people from traditionally underrepresented demographic and sociocultural groups in more narrowly obtained educational datasets. The second affordance is the ability to conduct large-scale field experiments at minimal cost. Researchers can quickly evaluate multiple theory-based interventions and draw causal conclusions about their efficacy in an authentic learning environment. Together, diverse big data and experimentation provide evidence on "what works for whom" that can extend theories to account for individual differences and support efforts to effectively target materials and support structures in online learning environments.

Keywords: Research methodology, big diverse data, randomized field experiments, inclusive science, MOOC

Massive open online courses (MOOCs) are a technological innovation for providing low-cost educational experiences to a worldwide audience. In 2012, some of the first MOOCs attracted hundreds of thousands of people from countries around the world (Waldrop, 2013), providing momentum for a disruption in higher education. Just a few years later, hundreds of institutions worldwide began to offer MOOCs on online learning platforms such as Coursera, EdX, and FutureLearn. Beyond expanding access to higher education, MOOCs have generated unprecedented amounts of educational data that have fuelled scholarship in existing academic communities and sparked interests in disciplines that were historically less involved in the learning sciences. This has amplified research in existing interdisciplinary communities and given rise to entirely new communities at the intersections of education, computer science, human factors, and statistics. For the field of learning analytics, we highlight two novel features of MOOCs that can enhance next-generation research: the availability of not just big but also diverse learner-level educational data, and the opportunity to run large online field experiments at low cost.

The first feature of MOOCs that supports innovative research is the amount and nature of the data that can be collected. MOOCs collect learner data that is deep as well as broad: fine-grained records from individual learners' interactions with content in the learning environment for a large number of learners (Thille et al., 2014). The dimensions of this recently available data enable applications of machine learning and data mining techniques that were previously infeasible. Beyond the large scale, however, the learner population is also considerably more diverse in MOOCs than in typical college courses. MOOCs attract more learners from countries that are not Western, Educated, Industrialized, Rich, and Democratic (WEIRD), the population on which most experimental social science is based (Henrich, Heine, & Norenzayan, 2010). Diverse data is critical for advancing inclusive scientific theories and educational practices, ones that apply to a broader population. Moreover, big diverse data enables research that identifies individual differences between demographic and sociocultural groups (e.g., heterogeneous effects of an intervention) that could not be investigated in existing research with small or homogenous samples.

The second feature of MOOCs that promises to increase the pace and impact of educational research is the ability to conduct online experiments rapidly, economically, and with high fidelity (Reich, 2015). In the technology industry, this is referred to as *A/B testing* to convey that individuals are randomly assigned to one of two experimental conditions. Online experimentation enables rapid iteration to test theories and practices, because multiple experiments can run in parallel and researchers can add, delete, and modify experiments in real time and at low cost. For instance, a researcher might compare different versions of lecture videos in a course (e.g., varying the introduction to a topic and how concepts are presented) and observe performance on subsequent assessments. Once enough data has been collected, the researcher may drop those lecture versions associated with the lowest scores and add new versions based on theory and the results of existing versions, and continue iterating. In this process, the researcher may find that a particular version works best for a specific group of learners, for instance, less educated learners. This may provide novel theoretical insights and it calls for adaptive presentation of content to optimize for learning. Discovery of individual differences and responsive adaptation of content is feasible in digital learning environments with large heterogeneous samples of learners, such as in MOOCs.

The goal of this chapter is to map out these two features of MOOCs, in light of the development of the field of learning analytics, and discuss how these features might advance the theory and practice of learning and instruction. We begin this chapter with a brief historical overview of the genesis and development of MOOC initiatives. We discuss the advantages of big data and diverse learner samples for research and we review work that has begun to leverage these affordances. Then we turn to the opportunities that open up through experimentation and rapid iteration, and discuss how these have been used thus far in MOOC platforms. We close the chapter with a discussion of current limitations and ways to leverage the opportunities of large-scale digital learning environments more effectively.

THE PAST AND PRESENT OF MOOCS

The development of MOOCs occurred in the context of a long tradition of efforts to increase access to education, including distance learning (e.g., correspondence schools, radio instruction), open access universities, and open educational resources (Simonson, Smaldino, Albright, & Zvacek, 2011). However, the notion of what constitutes a MOOC fundamentally shifted between 2008, when George Siemens and Stephen Downes facilitated the first MOOC (Siemens, 2013), and 2012, when the *New York Times* declared "the year of the MOOC" (Pappano, 2012). This shift led Siemens (2013) to distinguish the original cMOOCs, which are based on the connectivist pedagogical model that emphasizes collective knowledge creation without imposing a rigid course structure, from later xMOOCs (i.e., the MOOCs of 2012 and onwards), which are mostly based on the instructionist model of lecture-based courses with assessments and a rigid course structure. Stanford University Professors Sebastian Thrun, Daphne Koller, and Andrew Ng, who re-envisioned MOOCs as digital amplifications of their lecture classes to reach a broader audience, sparked this ideological shift. This vision led to the creation of several corporate and non-profit MOOC-providing organizations, most notably Coursera, Udacity, EdX, and FutureLearn. Institutions of higher education worldwide rushed to contribute to the growing number of courses, with each course attracting tens of thousands of learners (Waldrop, 2013).

The initial excitement and momentum began to fade once it became apparent that MOOCs fell short of delivering on the promise of providing universal low-cost higher education. The first sobering piece of evidence was that only a small percentage of learners who start a course go on to complete it (Clow, 2013; Breslow et al., 2013), and although completion is not everyone's goal (Kizilcec & Schneider, 2015; Kizilcec, Piech, & Schneider, 2013), this pattern suggests that critical barriers have remained unaddressed. The second sobering realization concerned the promise of advancing access for historically underserved populations. Many MOOC learners are already highly educated (Emanuel, 2013). Moreover, learners in the United States tend to live in more affluent areas and individuals with greater socioeconomic resources are more likely to earn a certificate (Hansen & Reich, 2015). Further evidence shows that socioeconomic achievement gaps in MOOCs occur worldwide in terms of education levels and levels of national development (Kizilcec, Saltarelli, Reich, & Cohen, 2017), and, moreover, that women underperform relative to men (Kizilcec & Halawa, 2015). These patterns may be partly due to structural, cultural, and educational barriers (e.g., Internet access, prior knowledge, language skills, culture-specific teaching methods). Additionally, learners may face social psychological barriers, such as the fear of being seen as less capable because of their social group (i.e., social identity threat) and feeling uncertain about their belonging in MOOCs from elite Western institutions (Kizilcec et al., 2017; Steele, Spencer, & Joshua, 2002; Walton & Cohen, 2007). At least in terms of completion rates, North American MOOCs have disproportionately benefited more privileged learners, posing a critical challenge for a technology designed to promote educational equity. This highlights the amplifying power of technology, namely that new technologies tend to reflect existing inequalities unless active steps are taken to address them. In fact, as evidence of a lack of support for underserved populations emerged, platform providers broadened their initial focus on prominent US partners and started pursuing international university partners, NGOs, and foreign governments. While early MOOC platforms focused on providing desktop-based learning experiences, platform development efforts also shifted towards expanding support for mobile devices as a way to increase access in developing countries, where mobile Internet is pervasive.

The issue of accreditation and certification for MOOC learning activities has been under constant consideration as the movement has matured. While content was originally provided free of charge, the value of a certificate (and a "verified certificate," where an individual's identity is linked more closely to their course activity) has attracted learners willing to pay for access to content or just to receive a certificate at the end. The certificate credential has also evolved over time. A number of institutions offer degrees independent of academic institutions (e.g., Udacity Nanodegrees); use MOOCs as a gateway opportunity to complete liberal arts courses online (e.g., the Arizona State University and EdX Freshman Academy partnership); create compact online postgraduate programs (e.g., MIT Microdegrees) with the option to use this credential as a pathway to a traditional graduate degree; and offer full graduate programs online (e.g., University of Illinois' iMBA and Data Science programs on the Coursera platform). There is an increasing supply of courses and short programs from various institutions, especially on popular topics like data science. Going forward, as more efficient marketplaces develop that better connect employers and educational institutions, we expect to see increased competition between educational institutions both to attract learners to their courses and to offer courses that can demonstrate superior workplace performance and career opportunities.

ENRICHING THEORY AND PRACTICE WITH DIVERSE BIG DATA IN EDUCATION

Theories of learning and instruction describe parts of a complex system (Mitchell, 2009). Any research that examines an instructional method or a learning strategy is therefore limited by its particular context, for example participants' prior knowledge or the subject area. Which of the hundreds of potential contextual attributes matter in a particular instance is difficult to predict and infeasible to test. We therefore rely on scientific theory to constrain this complexity and identify the variables that matter (Koedinger, Booth, & Klahr, 2013). Nonetheless, educational theory is never conclusive or all encompassing. Empirical research in education, and the social sciences at large, tends to focus on specific contexts to reduce complexity at the expense of external validity. In particular, much empirical research in the social sciences is based on studies of people in WEIRD contexts, such as US college students who participate in psychology lab studies (Henrich, Heine, & Norenzayan, 2010). This raises questions about the generalizability of existing results and models to different contexts and populations. These concerns have also been raised specifically about research on technology-enhanced education (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014; Blanchard, 2012). To address this challenge, researchers require access to learner samples that are larger and more diverse than traditionally available. Such diverse learner samples are commonplace in MOOCs.

The supply and range of courses available on MOOC platforms has steadily increased since their initial offerings (Shah, 2015). These courses have been created by institutions around the world, including universities, museums, and national institutes. In early 2016, Coursera announced that they had reached 18 million learners worldwide.[1] Most learners are located in the United States, China, India, and Brazil, with the strongest growth in enrollment in Mexico, Colombia, Brazil, and Russia. Four in ten learners are women on average, but the gender ratio ranges from 22% in Nigeria to 55% in the Philippines. Likewise, interest in various course topics varies by gender and location: business courses are most popular in France, while Polish learners prefer computer science, the subject area with (globally) the least gender balance. Coursera learners tend to be well educated: around 80% had already earned a bachelor's degree, according to a survey in 2015 (Zhenghao et al., 2015). This pattern resembles that of FutureLearn, a MOOC platform based in the United Kingdom, used by 3 million learners in 2016[2]: 73% hold degrees and, in contrast to Coursera, 62% are women. EdX[3] and Udacity,[4] two other major MOOC providers, had served six million and two million learners by 2016, respectively. Many other institutions offer MOOCs, either through traditional learning management systems (e.g., the Canvas Network), via institutionally deployed open-source platforms (e.g.,

[1] https://blog.coursera.org/post/142363925112
[2] https://about.futurelearn.com/press-releases/futurelearn-has-3-million-learners/
[3] http://blog.edx.org/edx-year-in-review?track=blog
[4] https://www.udacity.com/success

Open EdX), or through proprietary or custom developed platforms. Together, according to data collected by Class Central (Shah, 2015), 550 institutions have created 4,200 courses spanning virtually all disciplines that are reaching a remarkably heterogeneous population of over 35 million people worldwide.

While the data collected within a MOOC is high in velocity and volume, two of the three characteristics of *big data* (Laney, 2001), it can be limited with respect to variety, unless active measures are taken to achieve variety. Traditional educational systems collect both detailed demographic information (e.g., gender, ethnicity, proxies for socioeconomic status) and prior knowledge measures (e.g., prior college enrollments, high school grades, standardized test scores). However, these variables are not collected automatically in MOOCs in order to maintain a low barrier to entry. Many MOOC-providing institutions thus began to collect this data through optional surveys. A self-selected group of learners, who tend to be more committed to completing the course, are also those who tend to complete these surveys. Reich (2014) suggested that roughly one quarter of enrolled learners fill out a course survey. The sheer volume of collected survey responses can be high (often tens of thousands), though it is important to remember that these data represent a skewed sample of generally more motivated learners. Improved mechanisms for data collection that overcome current limitations for unobtrusively obtaining comprehensive information on learner backgrounds are needed. Nevertheless, currently available survey data supports the assumption that MOOC learners constitute a relatively heterogeneous population from around the globe.

Access to a heterogeneous learner population brings two major advantages for advancing educational theory and practice. First, when evaluating an instructional method or analytic model on a heterogeneous sample, the results are more representative of a diverse set of learners, which reduces the likelihood of drawing conclusions that have adverse consequences for underrepresented groups. Extending theory based on evidence from heterogeneous samples also promotes the development of more inclusive environments that support learners of various backgrounds. The second major advantage of heterogeneous learner samples is that they can reveal individual differences. Diversity is an essential ingredient for advancing an understanding of what works for whom and why — insights that enable effective tailoring of course materials and instructional methods. There is substantial room for improvement beyond tailoring to the “average learner,” who may not even resemble any of the actual learners (Rose, 2016). In fact, learning scientists have identified numerous variables that influence the efficacy of instructional methods, including prior knowledge, cognitive control, mental ability, and personality (Jonassen & Grabowski, 1993). For example, prior knowledge is a well-documented individual difference (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010), such that instructional methods that are relatively effective for novice learners can become ineffective, even counterproductive, for learners with increasing domain knowledge — a phenomenon known as expertise reversal (Kalyuga, Ayres, Chandler, & Sweller, 2003). Taken together, diverse big data can advance a more inclusive science that moves beyond tailoring to averages.

Although researchers have examined countless individual differences, there is a scarcity of replication studies in the field of education. Replications account for only 0.13% of published papers in the 100 major journals (Makel & Plucker, 2014), and many of these studies rely on relatively homogenous student populations in WEIRD countries. Even if study samples were more diverse, the meta-analysis of individual studies across different learning contexts would be complicated by variation in instructional conditions, much of which remains unobserved and therefore unaccounted for (Gašević, Dawson, Rogers, & Gašević, 2016). MOOCs and online learning environments more generally can begin to address this pressing issue. These environments are particularly well suited for conducting large-scale studies with diverse samples in an authentic learning context, and it is substantially faster and cheaper to conduct an exact replication study in a MOOC by rerunning the same course or by embedding the same study in another course. In fact, MOOCs are especially suitable for conducting disciplinary research into what instructional approaches are most effective for different groups of learners. For example, in physics education, different approaches to teaching the second law of thermodynamics have been proposed and tested (e.g., Cochran & Heron, 2006), but it is unclear which approach is most effective for learners from different parts of the world.

A critical dimension of diversity in MOOCs is geography, and thus culture. MOOCs assemble learners from Western and Eastern countries with their distinct cultural foundations of learning (Li, 2012). In Eastern countries (e.g., China, Japan), learning tends to be viewed as a virtuous, life-long process of self-perfection, based on Confucian influences, whereas in Western countries (e.g., US, Canada), learning is seen as a form of inquiry that serves the goal of understanding the world around us, based on Socratic and Baconian influences. Indeed, students from Confucian Asia who attend Western universities go through a period of academic adjustment (Rienties & Tempelaar, 2013) and this culture shock can hinder learning and raise feelings of alienation (Zhou, Jindal-Snape, Topping, &

Todman, 2008). This highlights the potential influence of cultural differences, and individual differences more broadly, on the efficacy of instructional approaches. MOOCs provide sufficiently diverse learner samples to investigate individual differences and refine existing theories by evaluating understudied dimensions of learner characteristics, both at an individual and group level. New insights into individual differences can inform current practices that support academic adjustment and tailored learning experiences in both in-person and online environments.

Current Research that Leverages Diverse Big Data in MOOCs

Researchers are just beginning to leverage the potential of large, heterogeneous learner data from MOOCs. Recent studies have investigated demographic and geographic differences in course navigation (Guo & Reinecke, 2014), learner motivation (Kizilcec & Schneider, 2015), persistence and achievement (DeBoer, Stump, Seaton, & Breslow, 2013; Kizilcec & Halawa, 2015; Kizilcec et al., 2013), and socioeconomic differences in course completion of learners in the United States (Hansen & Reich, 2015) and worldwide (Kizilcec et al., 2017). Notably, although learner demographics account for significant differences in course outcomes, they provide limited improvements over behavioural log data in the context of predictive modelling (Brooks, Thompson, & Teasley, 2015a; Brooks, Thompson, & Teasley, 2015b). We briefly describe two examples from the literature that highlight a range of possible approaches for leveraging diversity in MOOCs.

We first consider work by Kulkarni, Cambre, Kotturi, Bernstein, & Klemmer (2015), who set out to harness the diversity of MOOC learners to improve engagement and learning. They identified the relatively low level of social interaction in MOOCs as an opportunity for innovation and research. To address this shortcoming, they engineered a peer discussion system that puts online learners in touch with others in the course via group video chats. A series of experiments were conducted to examine the influence of group composition on performance on assessments and to refine the design of the peer discussion system. In three experiments, learners were assigned into discussion groups with high versus low geographical diversity in terms of how many countries were represented. Different outcome measures of performance were assessed in each experiment, encompassing an open-ended question to assess conceptual understanding of the session, scores on weekly "homework" assessments, and the final exam score. Confirming the authors' hypothesis, high-diversity peer discussions yielded short-term improvements in performance. Gender diversity, in contrast, showed no effect overall or differentially by diversity condition, counter to what might be predicted by prior work (Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). Their research demonstrates a promising avenue for leveraging diversity as an educational asset. It also begins to examine individual differences that may influence the efficacy of this approach by testing for gender effects and evaluating alternative operationalizations of diversity.

A second example from the literature concerns the optimal presentation of the instructor in lecture videos. Seeing a human face can make it easier to pay attention, but it can also be distracting. The image principle posits that showing the instructor in a video does not affect learning outcomes, because the motivational benefits of social cues are counterbalanced by additional extraneous cognitive processing (Mayer, 2001). How does this finding translate to the context of MOOCs, where motivation is a critical antecedent of persistence and achievement? Kizilcec, Bailenson, and Gomez (2015) found that 35% of MOOC learners in a course preferred watching videos without the face when given a choice, mainly because they found the face too distracting. Then, in a randomized experiment in the same MOOC, the default video that constantly showed the instructor was compared to a strategic version that omitted the instructor when it was distracting. The strategic presentation raised perceived cognitive load and social presence, but it had no overall effect on persistence or course grades. However, accounting for learning preference (i.e., whether individuals preferred learning from pictures and diagrams or from written and verbal information), there was a substantial individual difference in persistence: learners who expressed a verbal learning preference were 46% more likely to drop out of the course with the strategic than the constant presentation. This demonstrates the importance of both accounting for individual differences in practice and refining existing theories. If social cues are more distracting or motivating for different people, this insight is worth incorporating in learner models for targeted instructional design.

TESTING THEORY AND EVALUATING EDUCATIONAL PRACTICES USING ONLINE FIELD EXPERIMENTS

Compared to traditional learning management systems used in higher education, MOOCs offer a limited set of options to course designers and hardly any new features. However, the large and diverse learning community behind MOOCs provides a remarkable opportunity to learn more about learning and teaching through experimental research. Much of the initial research with MOOCs focused on the analysis of course log data collected by default (e.g., clickstreams) and self-report measures from course surveys with relatively low response rates. The recent availability of instructor-facing experimentation features in MOOCs has enabled researchers to conduct simple randomized experiments. Here we review three streams of experimental research — one concerned with small encouragements to promote engagement and learning, another concerned with changes to the course content and structure, and a third where MOOCs serve as a laboratory for studying general phenomena — and discuss methodological considerations going forward.

Three Streams of Published Experimental Research in MOOCs

One stream of experimental research has focused on small encouragements or nudges to improve course outcomes. These types of interventions can be conducted through email, for example, by randomly assigning learners to receive different messages. A number of studies employed A/B tests to increase participation in discussion forums. Lamb, Smilack, Ho, and Reich (2015) tested three treatments (a self-test participation check, discussion priming with summaries of prior discussions, and discussion preview emails about upcoming discussion topics) and found that the participation check increased forum activity over the default control condition. Kizilcec, Schneider, Cohen, and McFarland (2014) tested framing effects in email encouragements for forum participation in two experiments and found that a collectivist framing (i.e., "learn together", "help each other") reduced participation relative to an individualistic or neutral framing. Martinez (2014) tested framing effects using a social comparison paradigm (Festinger, 1954). Learners received an email with either an upward social comparison (describing how many learners outperform you), a downward social comparison (describing how many perform worse), or a control message omitting any social comparison. While the downward comparison motivated high-performing learners, struggling learners benefited from the upward comparison. Finally, Renz, Hoffmann, Staubitz, and Meinel (2016) found that emails exhibiting popular forum discussions and unanswered questions increased forum activity, and that reminder emails about unseen lecture videos increased course activity (i.e., lecture views), compared to other reminders. However, a downside of email interventions is that researchers typically cannot observe who opened the email and was exposed to the treatment, which raises an analytic challenge for estimating treatment effects (cf. Lamb et al., 2015). Survey experiments, with the experiment embedded inside a survey, offer an alternative. One study had survey-takers randomly assigned to receive either tips about self-regulated learning or a control message about course topics, but found no improvement in course outcomes (Kizilcec, Pérez-Sanagustín, & Maldonado, 2016). A potential downside of experiments in optional surveys is self-selection into the study, which tends to skew the sample towards more committed learners who may respond to the treatment differently from those who opted against taking the survey. In general, although small nudges can have surprisingly large effects on human behaviour (Thaler & Sunstein, 2009), most experiments in MOOCs have yielded small or non-significant results.

Another stream of experimental research has examined theory-based changes to course content and course structure. Renz, Hoffmann, Staubitz, and Meinel (2016) assessed the impact of providing learners with an "onboarding" session, an interactive tour that explains the course structure and navigation, but they found no improvements in course engagement. Following a quasi-experimental approach, Mullaney and Reich (2015) compared two consecutive instances of the same course with different content release models, staggered versus all-at-once presentation of materials. They also found no significant difference in persistence and completion rates. To facilitate two established learning strategies (retrieval practice and study planning), Davis, Chen, van der Zee, Hauff, and Houben (2016) tested weekly writing prompts that asked learners to summarize content and plan ahead. Yet, once again, no improvements in course persistence and completion were detected. Building on multimedia learning theory, Kizilcec and colleagues (2015) tested how the presentation of the instructor's face in video lectures influences attrition and achievement rates and they found heterogeneous effects on attrition, as previously described. In the context of discussion forums, Tomkin and Charlevoix (2014) tested the effect of instructor contact on various course outcomes. Their high-touch condition, which had instructors respond to forum questions and send weekly summaries, did not improve satisfaction, persistence, or completion rates, compared with a low-touch condition without instructor involvement. Coetzee, Fox, Hearst, and Hartmann (2014) evaluated the impact of adopting a reputation system in the discussion forum and found that it increased response times and the number of responses per post, but it had no effect on grades or persistence. Another study evaluated different reputation systems and found that a forum badging system that emphasized badge progress and upcoming badges increased forum activity (Anderson, Huttenlocher, Kleinberg, & Leskovec, 2014). Most studies in this stream of research, despite using stronger manipulations, also found no significant improvements in learning outcomes.

The third stream of experimental research leverages MOOCs as a lab environment to test general theories in a real-world context. For example, to test for (potentially unconscious) bias in online classes, Baker, Dee, Evans, and John (2015) planted messages in discussion forums across 126 MOOCs (1,008 messages total, eight per course) and randomly assigned learner names to be evocative of different races and genders. They found evidence of discrimination: instructors wrote more replies for white male names than for white female, Indian, and Chinese names. In testing social-psychological barriers to achievement, Kizilcec, Saltarelli, Reich, and Cohen (2017) found that theory-based intervention activities, designed to mitigate concerns about not belonging in the course, can effectively close the global achievement gap between learners in more versus less developed countries. On the benefits of diversity, Kulkarni et al. (2015) tested the role of geographical diversity in peer video discussion and found that being in a more diverse group improved subsequent test performance. Leveraging a natural experiment in peer assessment, Rogers and Feller (2016) found that exposure to exemplary peer performance causes attrition, due to the upward social comparison that undermines motivation and expected success. Again in the context of peer assessment, Kizilcec (2016) tested how the level of transparency about the peer grading process (i.e., how grades are adjusted and computed) affects learners' trust in peer grading. Results suggest that an explanation that highlights the fairness of the procedure can promote resilience in trust for learners who received a lower than expected grade. The studies in this stream of research focus on different phenomena in the context of MOOCs and their results hold promise for enriching theory and practice.

Methodological Considerations for Randomized Field Experiments in MOOCs

From this review of published research, it stands out that many experiments did not produce significant results. This may be surprising, given that MOOCs offer relatively large sample sizes that should render even practically insignificant differences statistically significant.[5] However, MOOC data exhibits substantial levels of variance in outcome measures (e.g., persistence, grades). While statistical power, the chance of detecting a true effect in data, increases with sample size, it decreases as data becomes noisier. When researchers underestimate the level of unexplained variance, it can result in underpowered studies that yield no significant findings. Yet this variance may actually signal the presence of individual differences that warrant further examination, for example, by testing for heterogeneous treatment effects. In general, when evaluating and reporting on experiments in MOOCs, it is advisable to focus on the magnitude of treatment effects in addition to their statistical significance. Researchers should explicitly separate planned confirmatory tests of hypotheses from ad-hoc exploratory analyses. Given the overwhelming number of possible outcomes and covariate measures in MOOC data, there is a real danger of increasing the Type I error rate (false positives) as a result of multiple testing, specification search, and researcher degrees of freedom (Gelman & Loken, 2013). To address this challenge, replication, pre-registration, and use of Bayesian alternatives to frequentist hypothesis testing (e.g., Kruschke, 2013) can help build robust scientific evidence going forward.

Despite existing within a phenomenon that is only four years old, randomized experiments in MOOCs are poised to deliver significant contributions to theory in education and related disciplines. Yet the promise of online field experiments for educational improvement has not been widely realized. Limits on the availability of real-time data and the level of access required to implement complex parallel experiments mean that most researchers are still only testing one idea at a time at the pace of new courses going live. A critical step towards rapid iteration with experiments in MOOCs is laying the groundwork for adaptive experimentation to accomplish the dual goal of learning through experimentation with the learner population and iteratively providing a better learning experience. This would provide a special opportunity for advancing scholarship in disciplinary teaching and learning. For example, instead of trying out a new way of teaching the concept of recursion and comparing test results with the previous cohort (a quasi-experimental design), multiple approaches to recursion can be taught simultaneously and their efficacy determined in short order. This would enable simultaneous tests of multiple educational theories in a domain and refining theory and practice by examining heterogeneous effects – a process that currently requires a whole community of researchers and substantial resources. Williams and colleagues (2014) proposed a first concept for adaptive experimentation in MOOCs in the form of MOOClets, which are small pieces of content that adapt based on results of ongoing experiments. Going forward, dynamic assignment to experimental conditions, for instance using a multi-armed bandit algorithm (Bather & Gittins, 1990), can enable rapid iteration over course designs, especially in combination with experimental system that supports complex and parallel designs, such as PlanOut (Bakshy, Eckles, & Bernstein, 2014). Overall, randomized field experiments in MOOCs offer researchers a novel opportunity to enrich theory and

[5] As is standard practice in the social sciences, the field of educational research has adopted the $p < 0.05$ criterion to evaluate the statistical significance of experimental results. Thus, if there is a less than 5% chance of obtaining an effect at least as extreme as in the sample data when the null hypothesis is true, then the null (e.g., equal condition means) is rejected.

practice at a fast pace.

CONCLUSION

The ability to deploy large randomized experiments rapidly to heterogeneous learner populations has the potential to disrupt the way educational research is carried out. While in the past learning theories might arise from the careful study of a small number of highly selective environments where control over conditions is difficult (e.g., higher education classroom studies in WEIRD contexts), it is now possible to deploy randomized experiments with high fidelity to tens of thousands of learners across the globe in a single course — an unprecedented opportunity in the field.

Traditional higher education research has faced two major pragmatic constraints to experimental inquiry. Perhaps the most significant constraint is the paradigm of responsibility of instruction. In the higher education classroom, the faculty member teaching the course tends to be completely responsible for the student experience. Faculty thus tend to take an equality-driven approach rather than an experimental approach, and ensure that all students within the class have equal access to support and interventions. Innovation in these circumstances tends to be through quasi-experimental methods, where learners in a given cohort or year of study are compared with those of other cohorts or years of study, introducing more confounding variables. In MOOCs, the paradigm is different, perhaps due in part to the broader constellation of actors, including institutional administration and vendor partners, who assume some responsibility for the success of learners. The culture of some of these actors around balancing risk and reward, especially within venture-capital funded enterprises where rapid prototyping and testing is the norm, has fostered more favorable attitudes towards experimental approaches to advance learner success.

A second constraint in traditional higher education research is the acceptable amount of risk and reward made available through experimentation. With hundreds of years of higher education demonstrating value to society, and tens or hundreds of thousands of dollars on the line in tuition for a given student, it is harder for researchers to make the ethical argument to engage in high-risk research. Yet in MOOC environments, the majority of learners enrol at no cost, and few are in peril of losing their livelihood over the results of a MOOC experiment gone awry.[6] This difference is reflected in institutional policy. Many institutions have strong protections for student records and privacy because of legal obligations such as FERPA in the United States. However, the same obligations do not exist for learners (or "users") in MOOCs, which lifts some of the constraints from policies concerning experimental research with online learners. This has two important ramifications and opportunities for researchers:

1. The population of MOOC learners is different and, in many ways, much more diverse than that of traditional educational research, in terms of learner demographics (age, race, cultural background, et cetera), prior knowledge, and motivations for taking courses. This broader representation, along with the vast numbers of learners, provides an opportunity for scholars to test the generalizability of learning theories across populations and to identity learning theories that are most applicable to specific groups of learners. This can enable quantitative approaches to problems that require such large datasets; for instance, Dillahunt, Ng, Fiesta, and Wang's (2016) study of low-income populations who use MOOCs for social mobility would be difficult to study quantitatively if not for the breadth of learners enrolled in MOOCs.
2. The ability to experiment directly in the learning platform at no cost enables researchers to leverage the volume and variance of learner data for greater scientific impact. This presents an opportunity to close the feedback loop in education by promoting the integration of research, theory, and practice. Given the breadth and depth of the learner population in MOOCs, there is a real possibility to build environments that (semi-)automatically adapt to the learner based on her experience, the experiences of other learners, and the underlying platform data.

In this chapter, we have called out two affordances of research with MOOCs — the availability of diverse big data and the ability to conduct randomized field experiments with rapid iteration — which we believe will enable a more inclusive and agile science of learning. The fields of learning analytics and educational data mining, characterized largely by their heavy adoption and investigation of computational methods, are well poised to answer this call and achieve even broader impact going forward.

[6] Recent approaches to accepting MOOC credentials as credit in higher education, such as through the Arizona State University Global Freshman Academy and the MIT Micro-Masters programs, have begun to change the stakes for learners.

REFERENCES

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: Seven research-based principles for smart teaching*. Jossey-Bass.

Anderson, A., Huttenlocher, D., Kleinberg, J., & Leskovec, J. (2014). Engaging with massive online courses. *Proceedings of the 23rd International Conference on World Wide Web* (WWW '14), 7–11 April 2014, Seoul, Republic of Korea (pp. 687–698). New York: ACM.

Baker, R., Dee, T., Evans, B., & John, J. (2015). Bias in online classes: Evidence from a field experiment. Paper presented at the SREE Spring 2015 Conference, *Learning Curves: Creating and Sustaining Gains from Early Childhood through Adulthood*, 5–7 March 2015, Washington, DC, USA.

Bakshy, E., Eckles, D., & Bernstein, M. S. (2014). Designing and deploying online field experiments. *Proceedings of the 23rd International Conference on World Wide Web* (WWW '14), 7–11 April 2014, Seoul, Republic of Korea (pp. 283–292). New York: ACM.

Bather, J. A., & Gittins, J. C. (1990). Multi-armed bandit allocation indices. *Journal of the Royal Statistical Society: Series A*, *153*(2), 257.

Blanchard, E. G. (2012). On the WEIRD nature of ITS/AIED conferences. In S. A. Cerri et al. (Eds.), *Proceedings of the 11th International Conference on Intelligent Tutoring Systems* (ITS 2012), 14–18 June 2012, Chania, Greece (pp. 280–285). Lecture Notes in Computer Science. Springer Berlin Heidelberg.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom research into edX's first MOOC. *Research & Practice in Assessment*, 8. http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF2.pdf.

Brooks, C., Thompson, C., & Teasley, S. (2015a). A time series interaction analysis method for building predictive models of learners using log data. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 126–135). New York: ACM.

Brooks, C., Thompson, C., & Teasley, S. (2015b). Who you are or what you do: Comparing the predictive power of demographics vs. activity patterns in massive open online courses (MOOCs). *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 245–248). New York: ACM.

Clow, D. (2013). MOOCs and the funnel of participation. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium. New York: ACM. doi:10.1145/2460296.2460332

Cochran, M. J., & Heron, P. R. L. (2006). Development and assessment of research-based tutorials on heat engines and the second law of thermodynamics. *American Journal of Physics*, 74(8), 734.

Coetzee, D., Fox, A., Hearst, M. A., & Hartmann, B. (2014). Should your MOOC forum use a reputation system? *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '14), 15–19 February 2014, Baltimore, Maryland, USA (pp. 1176–1187). New York: ACM.

Davis, D., Chen, G., van der Zee, T., Hauff, C., & Houben, G.-J. (2016). Retrieval practice and study planning in MOOCs: Exploring classroom-based self-regulated learning strategies at scale. *Proceedings of the 11th European Conference on Technology Enhanced Learning* (EC-TEL 2016), 13–16 September 2016, Lyon, France (pp. 57–71). Lecture Notes in Computer Science. Springer International Publishing.

DeBoer, J., Stump, G. S., Seaton, D., & Breslow, L. (2013). Diversity in MOOC students' backgrounds and behaviors in relationship to performance in 6.002x. *Proceedings of the 6th Conference of MIT's Learning International Networks Consortium* (LINC 2013), 16–19 June 2013, Cambridge, Massachusetts, USA. https://tll.mit.edu/sites/default/files/library/LINC%20'13.pdf

Dillahunt, T. R., Ng, S., Fiesta, M., & Wang, Z. (2016). Do massive open online course platforms support employability? *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '16), 27 February–2 March 2016, San Francisco, CA, USA (pp. 233–244). New York: ACM.

Emanuel, E. J. (2013). Online education: MOOCs taken by educated few. *Nature*, 503(7476), 342.

Festinger, L. (1954). A theory of social comparison processes. *Human Relations: Studies towards the Integration of the Social Sciences*, 7(2), 117–140.

Gašević, D., Dawson, S., Rogers, T., & Gašević, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28(January), 68–84.

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no "fishing expedition" or "p-hacking" and the research hypothesis was posited ahead of time. http://www.stat.columbia.edu/~gelman/research/unpublished/p_hacking.pdf.

Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. *Proceedings of the 1st ACM Conference on Learning @ Scale* (L@S 2014), 4–5 March 2014, Atlanta, Georgia, USA (pp. 21–30). New York: ACM.

Hansen, J. D., & Reich, J. (2015). Democratizing education? Examining access and usage patterns in massive open online courses. *Science*, 350(6265), 1245–1248.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, 33(2–3), 61–83; discussion 83–135.

Jonassen, D. H., & Grabowski, B. L. H. (1993). *Handbook of individual differences, learning, and instruction*. Routledge.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist*, 38(1), 23–31.

Kizilcec, R. F. (2016). How much information? Effects of transparency on trust in an algorithmic interface. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '16), 7–12 May 2016, San Jose, CA, USA (pp. 2390–2395). New York: ACM.

Kizilcec, R. F., Bailenson, J. N., & Gomez, C. J. (2015). The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology*, 107(3), 724–739.

Kizilcec, R. F., & Halawa, S. (2015). Attrition and achievement gaps in online learning. *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 57–66). New York: ACM.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2016). Recommending self-regulated learning strategies does not improve performance in a MOOC. *Proceedings of the 3rd ACM Conference on Learning @ Scale* (L@S 2016), 25–28 April 2016, Edinburgh, Scotland (pp. 101–104). New York: ACM.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium. New York: ACM.

Kizilcec, R. F., Saltarelli, A. J., Reich, J., & Cohen, G. L. (2017). Closing global achievement gaps in MOOCs. *Science*, 355(6322), 251-252.

Kizilcec, R. F., & Schneider, E. (2015). Motivation as a lens to understand online learners: Toward data-driven design with the OLEI scale. *ACM Transactions on Computer–Human Interaction*, 22(2), 1–24.

Kizilcec, R. F., Schneider, E., Cohen, G. L., & McFarland, D. A. (2014). Encouraging forum participation in online courses with collectivist, individualist and neutral motivational framings. *eLearning Papers*, 37, 13–22.

Koedinger, K. R., Booth, J. L., & Klahr, D. (2013). Instructional complexity and the science to constrain it. *Science*, 342(6161), 935–937.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t test. *Journal of Experimental Psychology: General*, 142(2), 573–603.

Kulkarni, C., Cambre, J., Kotturi, Y., Bernstein, M. S., & Klemmer, S. R. (2015). Talkabout: Making distance matter with small groups in massive classes. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15), 14–18 March 2015, Vancouver, BC, Canada (pp. 1116–1128). New York: ACM.

Lamb, A., Smilack, J., Ho, A., & Reich, J. (2015). Addressing common analytic challenges to randomized experiments in MOOCs: Attrition and zero-inflation. *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 21–30). New York: ACM.

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6, 70.

Li, J. (2012). *Cultural foundations of learning: East and west*. Cambridge University Press.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Educational Researcher*, 43(6), 304–316.

Martinez, I. (2014). MOOCs as a massive research laboratory. University of Virginia.

Mayer, R. E. (2001). *Multimedia learning*. Cambridge University Press.

Mitchell, S. D. (2009). *Unsimple truths: Science, complexity, and policy*. University of Chicago Press.

Mullaney, T., & Reich, J. (2015). Staggered versus all-at-once content release in massive open online courses: Evaluating a natural experiment. *Proceedings of the 2nd ACM Conference on Learning @ Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 185–194). New York: ACM.

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology*, 45(3), 487–501.

Pappano, L. (2012, November 2). The year of the MOOC. *The New York Times*. http://www.nytimes.com/2012/11/04/education/edlife/massive-open-online-courses-are-multiplying-at-a-rapid-pace.html

Reich, J. (2014, December 8). MOOC completion and retention in the context of student intent. *EDUCAUSE Review Online*. http://er.educause.edu/articles/2014/12/mooc-completion-and-retention-in-the-context-of-student-intent

Reich, J. (2015). Rebooting MOOC research. *Science*, 347(6217), 34–35.

Renz, J., Hoffmann, D., Staubitz, T., & Meinel, C. (2016). Using A/B testing in MOOC environments. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 304–313). New York: ACM.

Rienties, B., & Tempelaar, D. (2013). The role of cultural dimensions of international and Dutch students on academic and social integration and academic performance in the Netherlands. *International Journal of Intercultural Relations*, 37(2), 188–201.

Rogers, T., & Feller, A. (2016). Discouraged by peer excellence: Exposure to exemplary peer performance causes quitting. *Psychological Science*, 27(3), 365–374.

Rose, T. (2016). *The end of average: How we succeed in a world that values sameness*. HarperCollins.

Shah, D. (2015, December 28). MOOCs in 2015: Breaking down the numbers. *EdSurge*. https://www.edsurge.com/news/2015-12-28-moocs-in-2015-breaking-down-the-numbers

Siemens, G. (2013). Massive open online courses: Innovation in education? In R. McGreal, W. Kinuthia, & S. Marshall (Eds.), *Open educational resources: Innovation, research and practice* (pp. 5–16). Edmonton, AB: Athabasca University Press.

Simonson, M., Smaldino, S. E., Albright, M., & Zvacek, S. (2011). *Teaching and learning at a distance: Foundations of distance education*. Pearson Higher Ed.

Steele, C. M., Spencer, S. J., & Joshua, A. (2002). Contending with group image: The psychology of stereotype and social identity threat. *Advances in Experimental Social Psychology*, 34, 379–440.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Thille, C., Schneider, E., Kizilcec, R. F., Piech, C., Halawa, S. A., & Greene, D. K. (2014). The future of data-enriched assessment. *Research & Practice in Assessment*, 9, 5–16.

Tomkin, J. H., & Charlevoix, D. (2014). Do professors matter? Using an A/B test to evaluate the impact of instructor involvement on MOOC student outcomes. *Proceedings of the 1st ACM Conference on Learning @ Scale* (L@S 2014), 4–5 March 2014, Atlanta, Georgia, USA (pp. 71–78). New York: ACM.

Waldrop, M. M. (2013). Online learning: Campus 2.0. *Nature*, 495(7440), 160–163.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, 92(1), 82–96.

Williams, J. J., Li, N., Kim, J., Whitehill, J., Maldonado, S., Pechenizkiy, M., Chu, L., & Heffernan, N. (2014). The MOOClet framework: Improving online education through experimentation and personalization of modules. Working Paper No. 2523265, Social Science Research Network. doi:10.2139/ssrn.2523265

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688.

Zhenghao, C., Alcorn, B., Christensen, G., Eriksson, N., Koller, D., & Emanuel, E. J. (2015, September 22). Who's benefiting from MOOCs, and why. *Harvard Business Review*. https://hbr.org/2015/09/whos-benefiting-from-moocs-and-why

Zhou, Y., Jindal-Snape, D., Topping, K., & Todman, J. (2008). Theoretical models of culture shock and adaptation in international students in higher education. *Studies in Higher Education*, 33(1), 63–75.

Chapter 19: Predictive Modelling of Student Behavior Using Granular Large-Scale Action Data

Steven Tang, Joshua C. Peterson, and Zachary A. Pardos

Graduate School of Education, UC Berkeley, USA
DOI: 10.18608/hla17.019

ABSTRACT

Massive open online courses (MOOCs) generate a granular record of the actions learners choose to take as they interact with learning materials and complete exercises towards comprehension. With this high volume of sequential data and choice comes the potential to model student behaviour. There exist several methods for looking at longitudinal, sequential data like those recorded from learning environments. In the field of language modelling, traditional n-gram techniques and modern recurrent neural network (RNN) approaches have been applied to find structure in language algorithmically and predict the next word given the previous words in the sentence or paragraph as input. In this chapter, we draw an analogy to this work by treating student sequences of resource views and interactions in a MOOC as the inputs and predicting students' next interaction as outputs. Our approach learns the representation of resources in the MOOC without any explicit feature engineering required. This model could potentially be used to generate recommendations for which actions a student ought to take next to achieve success. Additionally, such a model automatically generates a student behavioural state, allowing for inference on performance and affect. Given that the MOOC used in our study had over 3,500 unique resources, predicting the exact resource that a student will interact with next might appear to be a difficult classification problem. We find that the syllabus (structure of the course) gives on average 23% accuracy in making this prediction, followed by the n-gram method with 70.4%, and RNN based methods with 72.2%. This research lays the groundwork for behaviour modelling of fine-grained time series student data using feature-engineering-free techniques.

Keywords: Behaviour modeling, sequence prediction, MOOCs, RNN

Today's digital world is marked with personalization based on massive logs of user actions. In the field of education, there continues to be research towards personalized and automated tutors that can tailor learning suggestions and outcomes to individual users based on the (often-latent) traits of the user. In recent years, higher education online learning environments such as massive open online courses (MOOCs) have collected high volumes of student-generated learning actions. In this chapter, we seek to contribute to the growing body of research that aims to utilize large sources of student-created data towards the ability to personalize learning pathways to make learning as accessible, robust, and efficient as desired. To do so, we demonstrate a strand of research focused on modelling the behavioural state of the student, distinct from research objectives concerned primarily with performance assessment and prediction. We seek to consider all actions of students in a MOOC, such as viewing lecture videos or replying to forum posts, and attempt to predict their next action. Such an approach makes use of the granular, non-assessment data collected in MOOCs and has potential to serve as a source of recommendations for students looking for navigational guidance.

Utilizing clickstream data across tens of thousands of students engaged in MOOCs, we ask whether generalizable patterns of actions across students navigating through MOOCs can be uncovered by modelling the behaviour of those who were ultimately successful in the course. Capturing the trends that successful students take through MOOCs can enable the development of automated recommendation systems so that struggling students can be given meaningful and effective recommendations to optimize their time spent trying to succeed. For this task, we utilize generative sequential models. Generative sequential models can take in a sequence of events as an input and generate a probability distribution over what event is likely to occur next. Two types of generative sequential models are utilized in this work, specifically the n-gram and the recurrent neural network (RNN) model, which have traditionally been successful when applied to other generative and sequential tasks.

This chapter specifically analyzes how well such models can predict the next action given a context of previous actions the student has taken in a MOOC. The purpose of such analysis would eventually be to create a system whereby an automated recommender could query the model to provide meaningful guidance on what action the student can take next. The next action in many cases may be the next resource prescribed by the course but in other cases, it may be a recommendation to consult a resource from a previous lesson or enrichment material buried in a corner of the courseware unknown to the student. These models we are training are known as generative, in that they can be used to generate what action could come next given a prior context of what actions the student has already taken. Actions can include things such as opening a lecture video, answering a quiz question, or navigating and replying to a forum post. This research serves as a foundation for applying sequential, generative models towards creating personalized recommenders in MOOCs with potential applications to other educational contexts with sequential data.

RELATED WORK

In the case of the English language, generative models are used to generate sample text or to evaluate the plausibility of a sample of text based on the model's understanding of how that language is structured. A simple but powerful model used in natural language processing (NLP) is the n-gram model (Brown, Desouza, Mercer, Pietra, & Lai, 1992), where a probability distribution is learned over every possible sequence of n terms from the training set. Recently, recurrent neural networks (RNNs) have been used to perform next-word prediction (Mikolov, Karafiát, Burget, Cernocky, & Khudanpur, 2010), where previously seen words are subsumed into a high dimensional continuous latent state. This latent state is a succinct numerical representation of all of the words previously seen in the context. The model can then utilize this representation to predict what words are likely to come next. Both of these generative models can be used to generate candidate sentences and words to complete sentences. In this work, rather than learning about the plausibility of sequences of words and sentences, the generative models will learn about the plausibility of sequences of actions undertaken by students in MOOC contexts. Then, such generative models can be used to generate recommendations for what the student ought to do next.

In the learning analytics community, there is related work where data generated by students, often in MOOC contexts, is analyzed. Analytics are performed with many different types of student-generated data, and there are many different types of prediction tasks. Crossley, Paquette, Dascalu, McNamara, and Baker (2016) provide an example of the paradigm where raw logs, in this case also from a MOOC, are summarized through a process of manual feature engineering. In our approach, feature representations are learned directly from the raw time series data. This approach does not require subject matter expertise to engineer features and is a potentially less lossy approach to utilizing the raw information in the MOOC clickstream. Pardos and Xu (2016) identified prior knowledge confounders to help improve the correlation of MOOC resource usage with knowledge acquisition. In that work, the presence of student self-selection is a source of noise and confounders. In contrast, student selection becomes the signal in behavioural modelling. In Reddy, Labutov, and Joachims (2016), multiple aspects of student learning in an online tutoring system are summarized together via embedding. This embedding process maps assignments, student ability, and lesson effectiveness onto a low dimensional space. Such a process allows for lesson and assignment pathways to be suggested based on the model's current estimate of student ability. The work in this chapter also seeks to suggest learning pathways for students, but differs in that additional student behaviours, such as forum post accesses and lecture video viewings, are also included in the model. Additionally, different generative models are employed.

In this chapter, we are working exclusively with event log data from MOOCs. While this user clickstream traverses many areas of interaction, examples of behaviour research have analyzed the content of the resources involved in these interaction sequences. Such examples include analyzing frames of MOOC videos to characterize the video's engagement level (Sharma, Biswas, Gandhi, Patil, & Deshmukh, 2016), analyzing the content of forum posts (Wen, Yang, &

Rosé, 2014; Reich, Stewart, Mavon, & Tingley, 2016), and analyzing the ad-hoc social networks that arise from interactions in the forums (Oleksandra & Shane, 2016). We are looking at all categories of possible student events at a more abstract level compared to these content-focused approaches.

In terms of cognition in learning analytics and EDM, much work has been done to assess the latent knowledge of students through models such as Bayesian knowledge tracing (BKT; Corbett & Anderson, 1994), including retrofitting the model to a MOOC (Pardos, Bergner, Seaton, & Pritchard, 2013) using superficial course structure as a source of knowledge components. This type of modelling views the actions of students as learning opportunities to model student latent knowledge. Student knowledge is not explicitly modelled in this chapter, though the work is related. Instead, our models focus on predicting the complement of this performance data, which is the behavioural data of the student.

Deep knowledge tracing (DKT; Piech et al., 2015) uses recurrent neural networks to create a continuous latent representation of students based on previously seen assessment results as they navigate online learning environments. In that work, recurrent neural networks summarize all of a student's prior assessment results by keeping track of a complex latent state. That work shows that a deep learning approach can be used to represent student knowledge, with favourable accuracy predictions relative to shallow BKT. Such results, however, are hypothesized to be explained by already existing extensions of BKT (Khajah, Lindsey, & Mozer, 2016). The use of deep learning to approach knowledge tracing still finds useful relationships in the data automatically, but potentially does not find additional representations relative to already proposed extensions to BKT. The work in this chapter is related to the use of deep networks to represent students, but differs in that all types of student actions are considered rather than only the use of assessment actions.

Specifically, in this chapter we consider using both the n-gram approach and a variant of the RNN known as the long short-term memory (LSTM) architecture (Hochreiter & Schmidhuber, 1997). These two both model sequences of data and provide a probability distribution of what token should come next. The use of LSTM architectures and similar variants have recently achieved impressive results in a variety of fields involving sequential data, including speech, image, and text analysis (Graves, Mohamed, & Hinton, 2013; Vinyals, Kaiser, et al., 2015; Vinyals, Toshev, Bengio, & Erhan, 2015), in part due to its mutable memory that allows for the capture of long- and short-range dependencies in sequences. Since student learning behaviour can be represented as a sequence of actions from a fixed action state space, LSTMs could potentially be used to capture complex patterns that characterize successful learning. In previous work, modelling of student clickstream data has shown promise with methods such as n-gram models (Wen & Rosé, 2014).

DATASET

The dataset used in this chapter came from a Statistics BerkeleyX MOOC from Spring 2013. The MOOC ran for five weeks, with video lectures, homework assignments, discussion forums, and two exams. The original dataset contains 17 million events from around 31,000 students, where each event is a record of a user interacting with the MOOC in some way. These interactions include events such as navigating to a particular URL in the course, up-voting a forum thread, answering a quiz question, and playing a lecture video. The data is processed so that each unique user has all of their events collected in sequential order: 3,687 types of events are possible. Every row in the dataset is converted to a particular index that represents the action taken or the URL accessed by the student.

Thus, every unique user's set of actions is represented by a sequence of indices, of which there are 3,687 unique kinds. Our recorded event history included students navigating to different pages of the course, which included forum threads, quizzes, video pages, and wiki pages. Within these pages, we also recorded the actions taken within the page, such as playing and pausing a video or checking a problem. We also record JavaScript navigations called sequential events. In this rendition of our pre-processing, we record these sequence events by themselves, without explicit association with the URL navigated to by the sequential event. Table 1 catalogs the different types of events present in the dataset as well as whether we chose to associate the specific URL tied to the event or not. In our pre-processing, some of these events are recorded as URL-specific, meaning that the model will be exposed to the exact URL the student is accessing for these events. Some events are recorded as non-URL-specific, meaning that the model will only know that the action took place, but not which URL that action is tied to in the course. Note that any event that occurred fewer than 40 times in the original dataset was filtered out. Thus, many of the forum events are filtered out, since they were URL-specific, but did not occur very frequently. Seq goto, seq next, and seq prev refer to events triggered when students select navigation buttons visible on the browser page. Seq next and seq prev will move to either the next or the previous content page in the course respectively, while a seq goto represents a jump within a section to any other section within a chapter.

Table 19.1. Logged Event Types and their Specificity

Course Page Events
Page View (URL-Specific)
Seq Goto (Non-URL-Specific)
Seq Next (Non-URL-Specific)
Seq Prev (Non-URL-Specific
Wiki Events
Page View (URL-Specific)
Video Events
Video Pause (Non-URL-Specific)
Video Play (Non-URL-Specific)
Problem Events
Problem View (URL-Specific)
Problem Check (Non-URL-Specific)
Problem Show Answer (Non-URL-specific)
Forum Events
Forum View (URL-Specific)
Forum Close (filtered out)
Forum Create (filtered out)
Forum Delete (filtered out)
Forum Endorse (filtered out)
Forum Follow (URL-Specific)
Forum Reply (URL-Specific)
Forum Search (Non-URL-specific)
Forum Un-follow (filtered out)
Forum Un-vote (filtered out)
Forum Update (filtered out)
Forum Up-vote (URL-Specific)
Forum View Followed Threads (URL-Specific)
Forum View Inline (URL-Specific)
Forum View User Profile (URL-Specific)

For example, if a student accesses the chapter 2, section 1 URL, plays a lecture video, clicks on the next arrow button (which performs a JavaScript navigation to access the next section), answers a quiz question, then clicks on section 5 within the navigation bar (which performs another JavaScript navigation), that student's sequence would be represented by five different indices. The first would correspond to the URL of chapter 2, section 1, the second to a play video token, the third to a navigation next event, the fourth to a unique identifier of which specific problem within the course the student accessed, and the fifth to a navigation goto event. The model would be given a list of these five indices in order, and trained to predict what should come after. The indices therefore represent the sequence of actions the student took. The length of five is not required; generative models can be given sequences of arbitrary length.

Of the 31,000 students, 8,094 completed enough assignments and scored high enough on the exams to be considered "certified" by the instructors of the course. Note that in other MOOC contexts, certification sometimes means that the student paid for a special certification, but that is not the case for this MOOC. The certified students accounted for 11.2 million of the original 17 million events, with an average of 1,390 events per certified student. The distinction between certified and non-certified is important for this chapter, as we chose to train the generative models only on actions from students considered "certified," under the hypothesis that the sequence of actions that certified students take might reasonably approximate a successful pattern of navigation for this MOOC.

Each row in the dataset contained relevant information about the action, such as the exact URL of what the user is accessing, a unique identifier for the user, the exact time the action occurs, and more. For this chapter, we do not consider time or other possibly relevant contextual information, but instead focus solely on the resource the student accesses or the action taken by the student. Events that occurred fewer than 40 times throughout the entire dataset were removed, as those tended to be rarely accessed discussion posts or user profile visits and are unlikely to be applicable to other students navigating through the MOOC.

METHODOLOGY

In this work, we investigate the use of two generative models, the recurrent neural network architecture, and the n-gram. In this section, we detail the architecture of the recurrent neural network and the LSTM extension, the model we hypothesize will perform best at next-action prediction. Other "shallow" models, such as the n-gram, are described afterwards.

Recurrent Neural Networks

Recurrent neural networks (RNNs) are a family of neural network models designed to handle arbitrary length sequential data. Recurrent neural networks work by keeping around a continuous, latent state that persists throughout the processing of a particular sequence. This latent state captures relevant information about the sequence so far, so that prediction at later parts of the sequence can be influenced by this continuous latent state. As the name implies, RNNs employ the computational approach utilized by feed forward neural networks while also imposing a recurring latent state that persists between time steps. Keeping the latent state around between elements in an input sequence is what gives recurrent neural networks their sequential modelling power. In this work, each input into the RNN will be a granular student event from the MOOC dataset. The RNN is trained to predict the student's next event based on the series of events seen so far. Figure 19.1 shows a diagram of a simple RNN, where inputs would be student actions and outputs would be the next student action from the sequence. The equations below show the mathematical operations used on each of the parameters of the RNN model: ht represents the continuous latent state. This latent state is kept around, such that the prediction at xt+1 is influenced by the latent state ht. The RNN model is parameterized by an input weight matrix Wx, recurrent weight matrix Wh, initial state h0, and output matrix Wy: bh and by are biases for latent and output units, respectively.

$$h_t = tanh(W_x x_t + W_h h_{t-1} + b_h) \tag{1}$$

$$y_t = \sigma(W_y h_t + b_y) \tag{2}$$

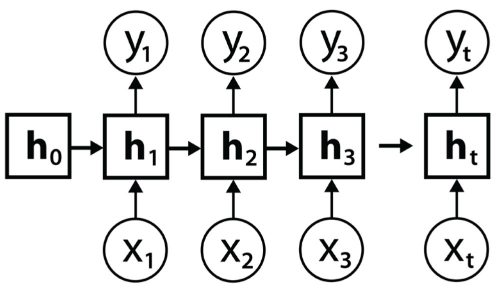


Figure 19.1. Simple recurrent neural network

LSTM Models

A popular variant of the RNN is the long short-term memory (LSTM; Hochreiter & Schmidhuber, 1997) architecture, which is thought to help RNNs learn long-range dependencies by the addition of "gates" that learn when to retain meaningful information in the latent state and when to clear or "forget" the latent state, allowing for meaningful long-term interactions to persist. LSTMs add additional gating parameters explicitly learned in order to determine when to clear and when to augment the latent state with useful information. Instead, each hidden state hi is replaced by an LSTM cell unit, which contains additional gating parameters. Because of these gates, LSTMs have been found to train more effectively than simple RNNs (Bengio, Simard, & Frasconi, 1994; Gers, Schmidhuber, & Cummins, 2000). The update equations for an LSTM are as follows:

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f) \tag{3}$$

$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i) \tag{4}$$

$$\tilde{C}_t = tanh(W_{Cx}x_t + W_{Ch}h_{t-1} + b_C) \tag{5}$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \tag{6}$$

$$o_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o) \tag{7}$$

$$h_t = o_t \times tanh(Ct) \tag{8}$$

Figure 19.2 illustrates the anatomy of a cell, where the numbers in the figure correspond to the previously mentioned update equations for the LSTM: ft, it, and ot represent the gating mechanisms used by the LSTM to determine "forgetting" data from the previous cell state, what to "in-put" into the new cell state, and what to output from the cell state. Ct represents the latent cell state for which information is removed from and added to as new inputs are fed into the LSTM. C˜t represents an intermediary new candidate cell state gated to update the next cell state.

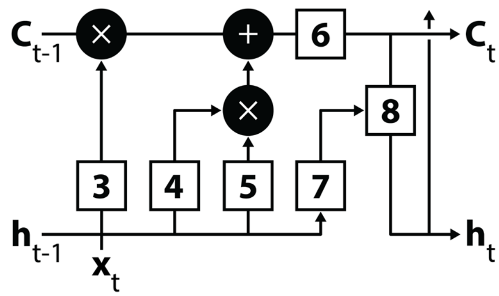


Figure 19.2. The anatomy of a cell with the numbers corresponding to the update equations for the LSTM.

LSTM Implementation

The generative LSTM models used in this chapter were implemented using Keras (Chollet, 2015), a Python library built on top of Theano (Bergstra et al., 2010; Bastien et al., 2012). The model takes each student action represented by an index number. These indices correspond to the index in a 1-hot encoding of vectors, also known as dummy variabilization. The model converts each index to an embedding vector, and then consumes the embedded vector one at a time. The use of an embedding layer is common in natural language processing and language modelling (Goldberg & Levy, 2014) as a way to map words to a multi-dimensional semantic space. An embedding layer is used here with the hypothesis that a similar mapping may occur for actions in the MOOC action space. The model is trained to predict the next student action, given actions previously taken by the student. Back propagation through time (Werbos, 1988) is used to train the LSTM parameters, using a softmax layer with the index of the next action as the ground truth. Categorical cross entropy is used calculating loss, and RMSprop is used as the optimizer. Drop out layers were added between LSTM layers as a method to curb overfitting (Pham, Bluche, Kermorvant, & Louradour, 2014). Drop out randomly zeros out a set percentage of network edge weights for each batch of training data. In future work, it may be worthwhile to evaluate other regularization techniques crafted specifically for LSTMs and RNNs (Zaremba, Sutskever, & Vinyals, 2014). We have made a version of our pre-processing and LSTM model code public,[1] which begins with extracting only the navigational actions from the dataset.

LSTM Hyperparameter Search

As part of our initial investigation, we trained 24 LSTM models each with a different set of hyperparameters for 10 epochs each. An epoch is the parameter-fitting algorithm making a full pass through the data. The searched space of hyperparameters for our LSTM models is shown in Table 19.2. These hyperparameters were chosen for grid search based on previous work that prioritized different hyperparameters based on effect size (Greff, Srivastava, Koutník, Steunebrink, & Schmidhuber, 2015). For the sake of time, we chose not to train 3-layer LSTM models with learning rates of .0001. We also performed an extended investigation, where we used the results from the initial investigation to serve as a starting point to explore additional hyperparameter and training methods.

Because training RNNs is relatively time consuming, the extended investigation consisted of a subset of promising hyperparameter combinations (see the Results section).

Table 19.2. LSTM Hyperparameter Grid

Hidden Layers	1	2	3
Nodes in Hidden Layer	64	128	256
Learning Rate (_)	0.01	0.001	.0001*

Cross Validation

To evaluate the predictive power of each model, 5-fold cross validation was used. Each model was trained on 80% of the data and then validated on the remaining 20%; this was done five times so that each set of student actions was in a validation set once. For the LSTMs, the model held out 10% of its training data to serve as the hill climbing set to provide information about validation accuracy during the training process. Each row in the held out set consists of the entire sequence of actions a student took. The proportion of correct next action predictions produced by the model is computed for each sequence of student actions. The proportions for an entire fold are averaged to generate the model's performance for that particular fold, and then the performances across all five folds are averaged to generate the CV-accuracy for a particular LSTM model hyperparameter set.

Shallow Models

N-*gram* models are simple, yet powerful probabilistic models that aim to capture the structure of sequences through the statistics of *n*-sized sub-sequences called *grams* and are equivalent to n-order Markov chains. Specifically, the model predicts each sequence state xi using the estimated conditional probability $P(x_i|x_{i-(n-1)}, \ldots, x_{i-1})$, which is the probability that xi follows the previous n-1 states in the training set. N-gram models are both fast and simple to compute, and have a straightforward interpretation. We expect *n*-grams to be an extremely competitive standard, as they are relatively high parameter models that essentially assign a parameter per possible action in the action space.

For the *n*-gram models, we evaluated those where *n* ranged from 2 to 10, the largest of which corresponds to the size of our LSTM context window during training. To handle predictions in which the training set contained no observations, we employed *backoff*, a method that recursively falls back on the prediction of the largest n-gram that contains at least one observation. Our validation strategy was identical to the LSTM models, wherein the average cross-validation score of the same five folds was computed for each model.

Course Structure Models

We also included a number of alternative models aimed at exploiting hypothesized structural characteristics of the sequence data. The first thing we noticed when inspecting the sequences was that certain actions are repeated several times in a row. For this reason, it is important to know how well this assumption alone predicts the next action in the dataset. Next, since course content is most often organized in a fixed sequence, we evaluated the ability of the course syllabus to predict the next page or action. We accomplished this by mapping course content pages to student page transitions in our action set, which yielded an overlap of 174 matches out of the total 300 items in the syllabus. Since we relied on matching content ID strings that were not always present in our action space, a small subset of possible overlapping actions were not mapped. Finally, we combined both models, wherein the current state was predicted as the next state if the current state was not in the syllabus.

RESULTS

In this section, we discuss the results from the previously mentioned LSTM models trained with different learning rates, number of hidden nodes per layer, and number of LSTM layers. Model success is determined through 5-fold cross validation and is related to how

[1] https://github.com/CAHLR/mooc-behavior-case-study

well the model predicts the next action. N-gram models, as well as other course structure models, are validated through 5-fold cross validation.

LSTM Models

Table 19.3 shows the CV-accuracy for all 24 LSTM models computed after 10 iterations of training. For the models with a learning rate of .01, accuracy on the hill climbing sets generally peaked at iteration 10. For the models with the lower learning rates, it would be reasonable to expect that peak CV-accuracies would improve through more training. We chose to simply report results after 10 iterations instead to provide a snapshot of how well these models are performing during the training process. We also hypothesize that model performance is unlikely to improve drastically over the .01 learning rate model performances in the long-run, and we need to maximize the most promising explorations to run on limited GPU computation resources. The best CV-accuracy for each learning rate is bolded for emphasis.

One downside of using LSTMs is that they require a GPU and are relatively slow to train. Thus, when investigating the best hyperparameters to use, we chose to train additional models based only on a subset of the initial explorations. We also extend the amount of context exposed to the model, extending past context from 10 elements to 100 elements. Table 4 shows these extended results. Each LSTM layer has 256 nodes and is trained for either 20 or 60 epochs, as opposed to just 10 epochs in the previous hyperparameter search results. The extended results show a large improvement over the previous results, where the new accuracy peaked at .7223 compared to .7093.

Figure 19.3 shows validation accuracy on the 10% hill-climbing hold out set during training by epoch for the 1 and 2 layer models from the initial exploration. Each data point represents the average hill-climbing accuracy among all three learning rates for a particular layer and node count combination. Empirically, having a higher number of nodes is associated with a higher accuracy in the first 10 epochs, while 2 layer models start with lower validation accuracies for a few epochs before approaching or surpassing the corresponding 1 layer model. This figure provides a snapshot for the first 10 epochs; clearly, for some parameter combinations, more epochs would result in a higher hill-climbing accuracy, as shown by the additional extended LSTM search. Extrapolating, 3-layer models may also follow the trend that the 2-layer models exhibited where accuracies may start lower initially before improving over their lower-layer counterparts.

Table 19.3. LSTM Performance (10 Epochs)

Learn Rate	Nodes	Layers	Accuracy
0.01	64	1	0.7014
0.01	64	2	0.7009
0.01	64	3	0.6997
0.01	128	1	0.7046
0.01	128	2	0.7064
0.01	128	3	0.7056
0.01	256	1	0.7073
0.01	256	2	**0.7093**
0.01	256	3	0.7092
0.001	64	1	0.6941
0.001	64	2	0.6968
0.001	64	3	0.6971
0.001	128	1	0.6994
0.001	128	2	0.7022
0.001	128	3	0.7026
0.001	256	1	0.7004
0.001	256	2	**0.7050**
0.001	256	3	**0.7050**
0.0001	64	1	0.6401
0.0001	64	2	0.4719
0.0001	128	1	0.6539
0.0001	128	2	0.6648
0.0001	256	1	0.6677
0.0001	256	2	**0.6894**

Table 19.4. Extended LSTM Performance (256 Nodes, 100 Window Size)

Learn Rate	Epoch	Layers	Accuracy
0.01	20	2	0.7190
0.01	60	2	0.7220
0.01	20	3	0.7174
0.01	60	3	**0.7223**
0.001	20	2	0.7044
0.001	60	2	0.7145
0.001	20	3	0.7039
0.001	60	3	**0.7147**

Course Structure Models

Model performance for the different course structure models is shown in Table 19.5. Results suggest that many actions can be predicted from simple heuristics such as stationarity (same as last), or course content structure. Combining both of these heuristics (“syllabus + repeat”) yields the best results, although none of the alternative models obtained performance within the range of the LSTM or n-gram results.

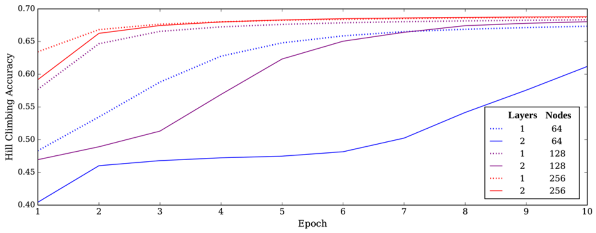


Figure 19.3. Average accuracy by epoch on hill climbing data, which comprised 10% of each training set.

Table 19.5. Structural Models

Structural Model	Accuracy
repeat	0.2908
syllabus	0.2339
syllabus + repeat	0.4533

N-gram Models

Model performance is shown in Table 19.6. The best performing models made predictions using either the previous 7 or 8 actions (8-gram and 9-gram respectively). Larger histories did not improve performance, indicating that our range of n was sufficiently large. Performance in general suggests that n-gram models were competitive with the LSTM models, although the best n-gram model performed worse than the best LSTM models. Table 19.7 shows the proportion of n-gram models used for the most complex model (10-gram). More than 62% of the predictions were made using 10-gram observations. Further, fewer than 1% of cases fell back on unigrams or bigrams to make predictions, suggesting that there was not a significant lack of observations for larger gram patterns.

Table 19.6. N-gram Performance

N-gram	Accuracy
2-gram	0.6304
3-gram	0.6658
4-gram	0.6893
5-gram	0.6969
6-gram	0.7012
7-gram	0.7030
8-gram	**0.7035**
9-gram	**0.7035**
10-gram	0.7033

Table 19.7. Proportion of 10-gram prediction by n

n	% Predicted by
1	0.0003
2	0.0084
3	0.0210
4	0.0423
5	0.0524
6	0.0605
7	0.0624
8	0.0615
9	0.0594
10	**0.6229**

Still, about 6% fewer data points looks to be predicted by successively larger n-grams.

Validating on Uncertified Students

We used the best performing "original" LSTM model after 10 epochs of training (.01 learn rate, 256 nodes, 2 layers) to predict actions on streams of data from students who did not ultimately end up certified. Many uncertified students only had a few logged actions, so we restricted analysis to students who had at least 30 logged actions. There were 10,761 students who met these criteria, with a total of 2,151,662 actions. The LSTM model was able to predict actions correctly from the uncertified student space with .6709 accuracy, compared to .7093 cross-validated accuracy for certified students. This difference shows that actions from certified students tend to be different than actions from uncertified students, perhaps showing potential application in providing an automated suggestion framework to help guide students.

Table 19.8. Cross Validated Models Comparison

N-gram	Correct	N-gram Incorrect
LSTM Correct	7,565,862	577,683
LSTM Incorrect	367,960	2,735,702

CONTRIBUTIONS

In this work, we approached the problem of modelling granular student action data by modelling all types of interactions within a MOOC. This differs in approach from previous work, which primarily focuses on modelling latent student knowledge using assessment results. In predicting a student's next action, the best performing LSTM model produced a cross-validation accuracy of .7223, which was an improvement over the best *n*-gram model accuracy of .7035: 210,000 more correct predictions of the total 11-million possible. Table 8 shows the number of times the two models agreed or disagreed on a correct or an incorrect prediction during cross validation. Both LSTM and *n*-gram models provide significant improvement over the structural model of predicting the next action by syllabus course structure and through repeats, which shows that patterns of student engagement clearly deviate from a completely linear navigation through the course material.

To our knowledge, this chapter marks the first time that behavioural data has been predicted at this level of granularity in a MOOC. It also represents the first time recurrent neural networks have been applied to MOOC data. We believe that this technique for representing students' behavioural states from raw time series data, without feature engineering, has broad applicability in any learning analytics context with high volume time series data. While our framing suggests how behavioural data models could be used to suggest future behaviours for students, the representation of their behavioural states could prove valuable for making a variety of other inferences on constructs ranging from performance to affect.

FUTURE WORK

Both the LSTM and the n-gram models have room for improvement. In particular, our n-gram models could benefit from a combination of backoff and smoothing techniques, which allow for better handling of unseen grams. Our LSTM may benefit from a broader hyper-parameter grid search, more training time, longer training context windows, and higher-dimensional action embeddings. Additionally, the signal-to-noise ratio in our dataset could be increased by removing less informative or redundant student actions, or adding additional tokens to represent time between actions.

The primary reason for applying deep learning models to large sets of student action data is to model student behaviour in MOOC settings, which leads to insights about how successful and unsuccessful students navigate through the course. These patterns can be leveraged to help in the creation of automated recommendation systems, wherein a struggling student can be provided with transition recommendations to view content based on their past behaviour and performance. To evaluate the possibility of such an application, we plan to test a recommendation system derived from our network against an undirected control group experimentally. Additionally, future work should assess performance of similar models for a variety of courses and examine to what extent course-general patterns can be learned using a single model. The models proposed in this chapter maintain a computational model of behaviour. It was demonstrated through these models that regularities do exist in student behaviour sequences in MOOCs. Given that a computational model was able to detect these patterns, what can the model tell us about student behaviours more broadly and how might those findings connect to and build upon existing behavioural theories? Since the model tracks a hidden behavioural state for the student at every time slice, this state can be visualized and correlated with other attributes of the students known to present at that time. Future work will seek to open up this computational model of behaviour so that it may help inform our own understanding of the student condition.

ACKNOWLEDGEMENTS

This work was supported by a grant from the National Science Foundation (IIS: BIGDATA 1547055).

REFERENCES

Bastien, F., Lamblin, P., Pascanu, R., Bergstra, J., Goodfellow, I. J., Bergeron, A., . . . Bengio, Y. (2012). Theano: New features and speed improvements. Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop. *Advances in Neural Information Processing Systems* 25 (NIPS 2012), 3–8 December 2012, Lake Tahoe, NV, USA. http://www.iro.umontreal.ca/~lisa/pointeurs/nips2012_deep_workshop_theano_final.pdf

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166.

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., . . . Bengio, Y. (2010, June). Theano: A CPU and GPU math expression compiler. *Proceedings of the Python for Scientific Computing Conference* (SciPy 2010), 28 June–3 July 2010, Austin, TX, USA (pp. 3–10).

Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D., & Lai, J. C. (1992). Class-based n-gram models of natural language. *Computational Linguistics*, 18(4), 467–479.

Chollet, F. (2015). Keras. GitHub. https://github.com/fchollet/keras

Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278.

Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 6–14). New York: ACM.

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471.

Goldberg, Y., & Levy, O. (2014). Word2vec explained: Deriving Mikolov et al.'s negative-sampling word-embedding method. CoRR. arxiv.org/abs/1402.3722

Graves, A., Mohamed, A.-r., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing* (ICASSP 2013), 26–31 May, Vancouver, BC, Canada (pp. 6645–6649). Institute of Electrical and Electronics Engineers.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2015). LSTM: A search space odyssey. arXiv preprint arXiv:1503.04069.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing? arXiv preprint arXiv:1604.02416.

Mikolov, T., Karafiát, M., Burget, L., Cernocky, J., & Khudanpur, S. (2010). Recurrent neural network based language model. *Proceedings of the 11th Annual Conference of the International Speech Communication Association* (INTERSPEECH 2010), 26–30 September 2010, Makuhari, Chiba, Japan (pp. 1045–1048). http://www.fit.vutbr.cz/research/groups/speech/publi/2010/mikolov_interspeech2010_IS100722.pdf

Oleksandra, P., & Shane, D. (2016). Untangling MOOC learner networks. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 208–212). New York: ACM.

Pardos, Z. A., Bergner, Y., Seaton, D. T., & Pritchard, D. E. (2013). Adapting Bayesian knowledge tracing to a massive open online course in EDX. In S. K. D'Mello et al. (Eds.), *Proceedings of the 6th International Conference on Educational Data Mining* (EDM2013), 6–9 July 2013, Memphis, TN, USA (pp. 137–144). International Educational Data Mining Society/Springer.

Pardos, Z. A., & Xu, Y. (2016). Improving efficacy attribution in a self-directed learning environment using prior knowledge individualization. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 435–439). New York: ACM.

Pham, V., Bluche, T., Kermorvant, C., & Louradour, J. (2014). Dropout improves recurrent neural networks for handwriting recognition. *Proceedings of the 14th International Conference on Frontiers in Handwriting Recognition* (ICFHR 2014) 1–4 September 2014, Crete, Greece (pp. 285–290).

Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L. J., & Sohl-Dickstein, J. (2015). Deep knowledge tracing. In C. Cortes et al. (Eds.), *Advances in Neural Information Processing Systems* 28 (NIPS 2015), 7–12 December 2015, Montreal, QC, Canada (pp. 505–513).

Reddy, S., Labutov, I., & Joachims, T. (2016). Latent skill embedding for personalized lesson sequence recommendation. CoRR. arxiv.org/abs/1602.07029

Reich, J., Stewart, B., Mavon, K., & Tingley, D. (2016). The civic mission of MOOCs: Measuring engagement across political differences in forums. *Proceedings of the 3rd ACM Conference on Learning @ Scale* (L@S 2016), 25–28 April 2016, Edinburgh, Scotland (pp. 1–10). New York: ACM.

Sharma, A., Biswas, A., Gandhi, A., Patil, S., & Deshmukh, O. (2016). Livelinet: A multimodal deep recurrent neural network to predict liveliness in educational videos. In T. Barnes et al. (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining* (EDM2016), 29 June–2 July 2016, Raleigh, NC, USA. International Educational Data Mining Society. http://www.educationaldatamining.org/EDM2016/proceedings/paper_64.pdf

Vinyals, O., Kaiser, L. Koo, T., Petrov, S., Sutskever, I., & Hinton, G. (2015). Grammar as a foreign language. In C. Cortes et al. (Eds.), *Advances in Neural Information Processing Systems* 28 (NIPS 2015), 7–12 December 2015, Montreal, QC, Canada (pp. 2755–2763). http://papers.nips.cc/paper/5635-grammar-as-a-foreign-language.pdf

Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015). Show and tell: A neural image caption generator. *Proceedings of the 2015 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR 2015), 8–10 June 2015, Boston, MA, USA. IEEE Computer Society. arXiv:1411.4555

Wen, M., & Rosé, C. P. (2014). Identifying latent study habits by mining learner behavior patterns in massive open online courses. *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management* (CIKM '14), 3–7 November 2014, Shanghai, China (pp. 1983–1986). New York: ACM.

Wen, M., Yang, D., & Rosé, C. P. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper et al. (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July 2014, London, UK. International Educational Data Mining Society. http://www.cs.cmu.edu/~mwen/papers/edm2014-camera-ready.pdf

Werbos, P. J. (1988). Generalization of backpropagation with application to a recurrent gas market model. *Neural Networks*, 1(4), 339–356.

Zaremba, W., Sutskever, I., & Vinyals, O. (2014). Recurrent neural network regularization. arXiv:1409.2329

Chapter 20: Applying Recommender Systems for Learning Analytics: A Tutorial

Soude Fazeli[1], Hendrik Drachsler[2], Peter Sloep[2]

[1] Welten Institute, Research Centre for Learning, Teaching and Technology, The Netherlands
[2] Open University of the Netherlands, The Netherlands
DOI: 10.18608/hla17.020

ABSTRACT

This chapter provides an example of how a recommender system experiment can be conducted in the domain of learning analytics (LA). The example study presented in this chapter followed a standard methodology for evaluating recommender systems in learning. The example is set in the context of the FP7 Open Discovery Space (ODS) project that aims to provide educational stakeholders in Europe with a social learning platform in a social network similar to Facebook, but unlike Facebook, exclusively for learning and knowledge sharing. In this chapter, we describe a full recommender system data study in a stepwise process. Furthermore, we outline shortcomings for data-driven studies in the domain of learning and emphasize the high need for an open learning analytics platform as suggested by the SoLAR society.

Keywords: Recommender system, offline data study, user study, collaborative filtering, information search and retrieval, information filtering, research, methodology, sparsity

With the emergence of massive amounts of data in various domains, recommender systems have become a practical approach to provide users with the most suitable information based on their past behaviour and current context. Duval (2011) introduced recommenders as a solution "to deal with the 'paradox of choice' and turn the abundance from a problem into an asset for learning" (p. 9), pointing out that several domains such as educational data mining, big data, and Web analytics all try to find patterns in large amounts of data. For instance, data mining approaches can make recommendations based on similarity patterns detected from the collected data of users. Furthermore, a survey conducted by Greller and Drachsler (2012), identified recommender systems and personalization as an important part of LA research.

Recommender systems can be differentiated according to their underlying technology and algorithms. Roughly, they are either content-based or use collaborative filtering. Content-based algorithms are one of the main methods used in recommender systems; they recommend an item to the user by comparing the representation of the item's content with the user's preference model (Pazzani & Billsus, 2007). Collaborative filtering is based on users' opinions and feedback on items. Collaborative filtering algorithms first find like-minded users and introduce them as so-called nearest neighbours to some target user; then they predict an item's rating for that user on the basis of the ratings given to this item by the target users' nearest neighbours (co-ratings) (Herlocker, Konstan, Terveen, & Riedl, 2004; Manouselis, Drachsler, Verbert, & Duval, 2012; Schafer, Frankowski, Herlocker, & Sen, 2007).

In the past, we have applied recommender systems in various educational projects with different objectives (Drachsler et al., 2010; Fazeli, Loni, Drachsler, & Sloep, 2014; Drachsler et al., 2009). In this chapter we want to share some best practices we have identified so far regarding the development and evaluation of recommender system algorithms in education; we especially want to provide an example of how to set up and run a recommender systems experiment.

As described by the RecSysTEL working group for Recommender Systems in Technology-Enhanced Learning (Drachsler, Verbert, Santos, & Manouselis, 2015) it is important to apply a standard evaluation method. The working group identified a research methodology consisting of four critical steps for evaluating a recommender system in education:

1. A *selection of dataset(s)* that suit the recommendation task. For instance, the recommendation task can be finding new items or finding relevant items for a user.
2. An *offline data study* of different algorithms on the selected datasets including well-known datasets (if possible, education-oriented datasets such as MovieLens makes movie recommendations) to provide insights into the performance of the recommender systems.
3. A comprehensive *user study* to test psycho-educational effects on learners as well as on the technical aspects of the designed recommender system.
4. A *deployment* of the recommender system in a real life application, where it can be tested under realistic, normal operational conditions with actual users.

The above four steps should be accompanied by a complete description of the recommender system according to the classification framework presented (Drachsler et al., 2015). The dataset used should be reported in the special section on educational datasets of the Journal of Learning Analytics[1] and made available for other researchers under certain conditions (Dietze, Siemens, Taibi, & Drachsler, 2016). This would allow other researchers to repeat and adjust any part of the research to gain comparable results and new insights and thus build up a body of knowledge around recommender systems in learning analytics.

In this chapter, we present an example of an experimental study that followed the research methodology described above for recommender systems in education. The rest of the chapter is organized as follows: In the next section, we present an example of a recommender system study that followed the methodology described above step by step. Next, we explain the practical implications of the experiment; then, we conclude.

A RECOMMENDER SYSTEM EXPERIMENT IN THE EDUCATIONAL DOMAIN

In this section, we describe how one should evaluate a recommender system in learning, making use of an experimental study presented in our 2014 EC-TEL paper (Fazeli et al., 2014). This study follows the standard methodology described above. To this methodology, however, we added an additional step: that of developing a conceptual/theoretical model (Fazeli et al., 2013)., which is presented in a RecSysTEL special issue (Manouselis et al., 2012).

In our study, our target environment is social learning platforms in general. Social learning platforms work similarly to social networks such as Facebook but, unlike Facebook, they are developed exclusively for the purpose of learning and knowledge sharing. They often serve, therefore, as a common place exclusively for educational stakeholders such as teachers, students, learners, policy makers, and so on. Our target social learning platform is Open Discovery Space (ODS).[2] As indicated on the ODS homepage,

> ODS addresses various challenges that face the eLearning environment in the European context. The interface has been designed with students, teachers, parents and policy makers in mind. ODS will fulfill three principal objectives. Firstly, it will empower stakeholders through a single, integrated access point for eLearning resources from dispersed educational repositories. Secondly, it engages stakeholders in the production of meaningful educational activities by using a social-network style multilingual portal, offering eLearning resources as well as services for the production of educational activities. Thirdly, it will assess the impact of the new educational activities, which could serve as a prototype to be adopted by stakeholders in school education.

The main goal of our study is to find out which recommender system can best suit the data and information needs of a social learning platform, the main recommendation task being to finding relevant items for users. In the following sub-sections, we describe the study step by step.

Dataset Selection

Most data studies target a specific environment or specific group of users and thus require a specific type of data. In our case, the target social learning platforms is ODS. Consequently, we tried to find data collected from learning platforms similar to ODS. We chose the MACE and OpenScout datasets for the following reasons:

1. The datasets provide social data of users (ratings, tags, reviews, et cetera) on learning resources. So, the structure, content, and target users of the datasets are similar to those of ODS.
2. Running recommender algorithms on these datasets helps us to evaluate their performance before going online with the actual users of the ODS.
3. Both the MACE and OpenScout datasets comply with the CAM (Context Automated Metadata) format (Schmitz et al., 2009), which offers a standard metadata specification for collecting and storing

[1] https://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/5071/5600

[2] http://opendiscoveryspace.eu

Table 20.1. Details of the Selected Datasets

Dataset	# of users	# of items	Transactions	Sparsity (%)	Source
MACE	631	12,571	23,032	99.70	MACE portal
OpenScout	331	1,568	2,560	99.50	OpenScout portal
MovieLens	941	1,512	96,719	93.69	GroupLens research

social data. CAM has also been applied in the ODS for storing social data.

Besides these two datasets, we also tested the MovieLens dataset as a reference since, up until now, the educational domain has been lacking reference datasets for study, unlike the ACM RecSys conference series, which deals with recommender systems in general. Table 20.1 provides an overview of all three datasets (Fazeli et al., 2014). Note that the educational datasets MACE and OpenScout clearly suffer from extreme sparsity. All the data are described more fully our EC-TEL 2014 article (Fazeli et al., 2014).

Offline Data Study

Algorithms. In this second step, we tried to select algorithms that would work well with our data. First, it is important to check the input data to be fed into the recommender algorithms. In this case, the ODS data, thus the data of the selected datasets, includes interaction data of users with learning resources (items). Therefore, we chose to use the Collaborative Filtering (CF) family of recommender systems. CF algorithms rely on the interaction data of users, such as ratings, bookmarks, views, likes, et cetera, rather than on the content data used by content-based recommenders. CF recommenders can be either memory-based or model-based, according to the "type"; they can be either item-based or user-based, referring to the "technique." For a detailed description of these distinctions, please see Section 4 of Fazeli et al. (2014). In our study, we made use of all types and techniques: both memory-based and model-based, as well as both user-based and item-based. Figure 20.1 shows our experimental method, consisting of three main steps:

1. We compared performance of memory-based CFs, including both user-based and item-based, by employing different similarity functions.
2. We ran the model-based CFs, including state-of-the-art Matrix Factorization methods, on our sample data.
3. We performed a final comparison of best-performing algorithms from steps 1 and 2. In addition to the baselines, we evaluated a graph-based approach proposed to enhance the mechanism of finding neighbours using the conventional k-nearest neighbours (kNN) method (Fazeli et al., 2014).

Performance Evaluation. After choosing suitable datasets and recommender algorithms, we arrive at the task of evaluating the performance of candidate algorithms. For this, we need to define an evaluation protocol (Herlocker et al., 2004). A good description of an evaluation protocol should address the following questions:

Q1. What is going to be measured?

Typically, in most offline recommender system studies, we measure the *prediction accuracy* of the recommendations generated. By this, we want to measure how much the rating predictions differ from the actual ones by comparing a training set and a test set. The training and test sets result from splitting our user ratings data (the same as user interaction data). In our

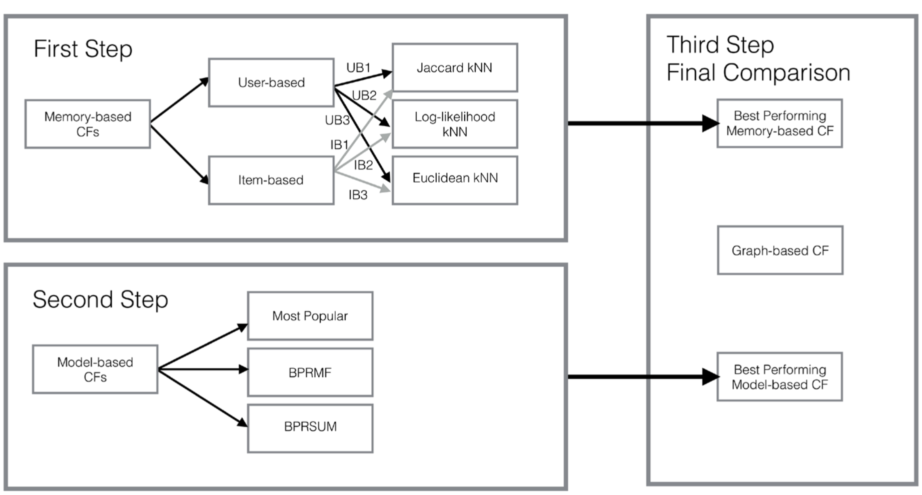


Figure 20.1. Experimental method used in Fazeli et al., 2014.

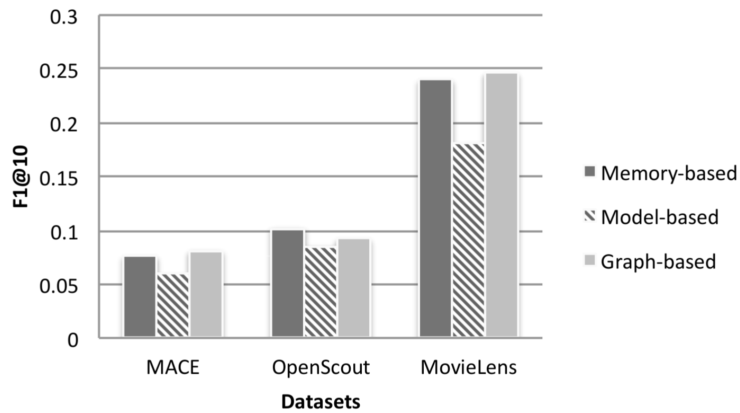


Figure 20.2. F1 of the graph-based CF and the best performing baseline memory-based and model-based CFs for all the datasets used (Fazeli et al., 2014).

EC-TEL 2014 study, we split user ratings into 80% and 20% for the training set and the test set, respectively. This kind of split is commonly used in recommender systems evaluations (Fazeli et al., 2014).

Q2. Which metrics are suitable for a recommender system study?

If our input data contains explicit user preferences, such as 5-star ratings, we can use MAE (mean average error) or RMSE (root mean square error). MAE and RMSE both follow the same range as the user ratings; for example, if the data contains 5-star ratings, these metrics range from 1 to 5.

If the input data contains implicit user preferences, such as views, bookmarks, downloads, et cetera, we can use Precision, Recall, and F1 scores. We made use of the F1 score since it combines precision and recall, which are both important metrics in evaluating the accuracy and coverage of the recommendations generated (Herlocker et al., 2004). F1 ranges from 0 to 1.

In addition, we need to define the n in top-n recommendations on which a metric is measured, also known as a *cut-off*. In Fazeli et al. (2014), we computed the F1 for the top 10 recommendations of the result set for each user.

Finally, we present the results of running the candidate algorithms on the datasets following the defined evaluation protocol. Due to limited space, we only present the final results of our EC-TEL 2014 article here. Please see Sections 5.1 and 5.2 of the original article (Fazeli et al., 2014) for more results.

Figure 20.2 shows the F1 results of best performing memory-based CF (Jaccard kNN), model-based CF (a Bayesian method), compared to the graph-based CF. The x-axis indicates the datasets used and the y-axis shows the values of F1. As Figure 20.2 shows, the graph-based approach performs best for MACE (8%) and MovieLens (24%) and the selected memory-based and model-based CFs come in second and third place right after the graph-based CF. For OpenScout, the memory-based approach performs better with a difference of almost 1%.

In conclusion, according to the results presented in Figure 20.2, the graph-based approach seems to perform well for social learning platforms. This is reflected by an improved F1, which is an effective combination of precision and recall of the recommendation made.

Deployment of the Recommender System and User Study

In the educational domain, the importance of user studies has become ever more apparent (Drachsler et al., 2015). Since the main aim of recommender systems in education goes beyond accurate predictions, it should extend to other quality indicators such as usefulness, novelty, and diversity of the recommendations. However, the majority of recommender system studies still rely on offline data studies alone. This is probably because user studies are time consuming and complicated.

After running the offline data study on the ODS data, we furthered the work reported in Fazeli et al. (2014) by conducting a user study with our target platform. For this, we integrated the algorithms that performed best with ODS. We asked actual users of ODS whether they were satisfied with the recommendations we made for them. For this we used a short questionnaire using five metrics: usefulness, accuracy, novelty, diversity, and serendipity. The full description and results of this data study and the follow-up user study have not been published yet. The user study does not confirm the results of the data study we had run on the actual ODS data, showing that it is quite necessary to run user studies that can go beyond the success indictors of data studies, such as prediction accuracy.

Accuracy is one of the important metrics in evaluating recommender systems but relying solely on this metric can lead data scientists and educational technologists down less effective pathways.

PRACTICAL IMPLICATIONS AND LIMITATIONS

Accessing most educational datasets is challenging since they are not publicly and openly available, for example through reference links. Moreover, it is often difficult to compare the findings of related studies, for instance those by Verbert et al. (2011) and Manouselis, Vuorikari, and Van Assche (2010). Although we applied the same datasets, and some of the algorithms used in those two studies, the results of our example study differ from their results. Therefore, we could not gain additional information from the comparisons regarding the personalization of learning resources. One possible reason is that the studies use different versions of the same dataset because the collected data belongs to different periods of time. For the MACE dataset, for instance, different versions are available. In fact, no unique version has been fixed for running experiments nor for comparison in the recommender system community.

This problem originates from the fact that, unfortunately, there is no gold-standard dataset in the educational domain comparable to the MovieLens dataset[3] in the e-commerce world. In fact, the LA community is in need of several representative datasets that can be used as a main set of references for different personalization approaches. The main aim is to achieve a standard data format to run LA research. This idea was initially suggested by the dataTEL project (Drachsler et al., 2011) and later followed up by the SoLAR Foundation for Learning Analytics (Gašević et al., 2011). In the domain of MOOCs, Drachsler and Kalz (2016) have discussed this lack of comparable results and the pressing need for a research cycle that uses data repositories to compare scientific results. Moreover, an EU-funded project called LinkedUp[4] follows a promising approach towards providing a set of gold-standard datasets by applying linked data concepts (Bizer, Heath, & Berners-Lee, 2009). The LinkedUp project aims to provide a linked data pool for learning analytics research and to run several data competitions through the central data pool.

Overall, the outcomes of different recommender systems or personalization approaches in the education domain are still hardly comparable due to the diversity of algorithms, learner models, datasets, and evaluation criteria (Drachsler et al., 2015; Manouselis et al., 2012).

CONCLUSION

The main goal of this chapter has been to illustrate how to identify the most appropriate recommender system for a learning environment. To do so, we followed an example data study using the standard methodology presented in Drachsler et al. (2015) for evaluating recommender systems in learning. The methodology consists of four main steps:

1. Select suitable datasets preferably from the educational domain and, in case the actual data is not available yet, similarly to the target data.
2. Run a set of candidate recommender algorithms that best fits the input data. The output of this step should reveal which recommender algorithms best works with the input data.
3. Conduct a user study to measure user satisfaction on the recommendations made for them.
4. Deploy the best candidate recommender to the target learning platform.

The fact that our user study results did not confirm the results of the offline data study illustrates the importance of running user studies even though they are quite time consuming and complicated.

ACKNOWLEDGMENTS

This chapter has been partly funded by the EU FP7 Open Discovery Space project. This document does not represent the opinions of the European Union, and the European Union is not responsible for any use that might be made of its content. The work of Hendrik Drachsler has been supported by the FP7 EU project LACE.

REFERENCES

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22.

Dietze, S., Siemens, G., Taibi, D., & Drachsler, H. (2016). Editorial: Datasets for learning analytics. *Journal of Learning Analytics*, 3(3), 307–311. http://dx.doi.org/10.18608/jla.2016.32.15

[3] http://www.grouplens.org/node/73
[4] www.linkedup-project.eu

Drachsler, H., Hummel, H., van den Berg, B., Eshuis, J., Waterink, W., Nadolski, R., Berlanga, A., Boers, N., & Koper, R. (2009). Effects of the ISIS recommender system for navigation support in self-organised learning networks. *Journal of Educational Technology & Society*, *12*(3), 115–126.

Drachsler, H., & Kalz, M. (2016). The MOOC and learning analytics innovation cycle (MOLAC): A reflective summary of ongoing research and its challenges. *Journal of Computer Assisted Learning*, 32(3), 281–290. http://doi.org/ 10.1111/jcal.12135

Drachsler, H., Rutledge, L., Van Rosmalen, P., Hummel, H., Pecceu, D., Arts, T., Hutten, E., & Koper, R. (2010). ReMashed: An usability study of a recommender system for mash-ups for learning. *International Journal of Emerging Technologies in Learning*, 5. http://online-journals.org/i-jet/article/view/1191

Drachsler, H., Verbert, K., Santos, O., & Manouselis, N. (2015). Panorama of recommender systems to support learning. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender systems handbook*, 2nd ed. Springer.

Drachsler, H., Verbert, K., Sicilia, M.-A., Wolpers, M., Manouselis, N., Vuorikari, R., Lindstaedt, S., & Fischer, F. (2011). *dataTEL: Datasets for Technology Enhanced Learning — White Paper*. Stellar Open Archive. http://dspace.ou.nl/handle/1820/3846

Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 9–17). New York: ACM.

Fazeli, S., Loni, B., Drachsler, H., & Sloep, P. (2014). Which recommender system can best fit social learning platforms? *Proceedings of the 9th European Conference on Technology Enhanced Learning* (EC-TEL '14), 16–19 September 2014, Graz, Austria (pp. 84–97).

Gašević, G., Dawson, C., Ferguson, S. B., Duval, E., Verbert, K., & Baker, R. S. J. d. (2011, July 28). *Open learning analytics: An integrated and modularized platform*. http://www.elearnspace.org/blog/wp-content/uploads/2016/02/ProposalLearningAnalyticsModel_SoLAR.pdf

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 42–57). New York: ACM.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1), 5–53.

Manouselis, N., Drachsler, H., Verbert, K., & Duval, E. (2012). *Recommender systems for learning*. Springer Berlin Heidelberg.

Manouselis, N., Vuorikari, R., & Van Assche, F. (2010). Collaborative recommendation of e-learning resources: An experimental investigation. *Journal of Computer Assisted Learning*, 26(4), 227–242.

Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web: Methods and strategies of web personalization* (pp. 325–341). Lecture Notes in Computer Science vol. 4321. Springer.

Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). Collaborative filtering recommender systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The adaptive web: Methods and strategies of web personalization* (pp. 291–324). Lecture Notes in Computer Science vol. 4321. Springer.

Schmitz, H., Scheffel, M., Friedrich, M., Jahn, M., Niemann, K., Wolpers, M., & Augustin, S. (2009). CAMera for PLE. In U. Cress, V. Dimitrova, & M. Specht (Eds.), *Learning in the Synergy of Multiple Disciplines: 4th European Conference on Technology Enhanced Learning (EC-TEL 2009) Nice, France, September 29–October 2, 2009 Proceedings* (pp. 507–520). Lecture Notes in Computer Science vol. 5794. Springer.

Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011). Dataset-driven research for improving recommender systems for learning. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 44–53). New York: ACM.

Chapter 21: Learning Analytics for Self-Regulated Learning

Philip H. Winne

Faculty of Education, Simon Fraser University, Canada
DOI: 10.18608/hla17.021

ABSTRACT

The Winne-Hadwin (1998) model of self-regulated learning (SRL), elaborated by Winne's (2011, in press) model of cognitive operations, provides a framework for conceptualizing key issues concerning kinds of data and analyses of data for generating learning analytics about SRL. Trace data are recommended as observable indicators that support valid inferences about a learner's metacognitive monitoring and metacognitive control that constitute SRL. Characteristics of instrumentation for gathering ambient trace data via software learners can use to carry out everyday studying are described. Critical issues are discussed regarding what to trace about SRL, attributes of instrumentation for gathering ambient trace data, computational issues arising when analyzing trace and complementary data, the scheduling and delivery of learning analytics, and kinds of information to convey in learning analytics that support productive SRL.

Keywords: Grain size, metacognition, self-regulated learning (SRL), traces

Four descriptions of learning analytics are widely cited. Siemens (2010) described learning analytics as "the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning." The website for the 1st International Conference on Learning Analytics and Knowledge posted this description: "the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs."[1] Educause (n.d.) defined learning analytics as "the use of data and models to predict student progress and performance, and the ability to act on that information." Building on Eckerson's (2006) framework, Elias (2011) notes "learning analytics seeks [sic] to capitalize on the modelling capacity of analytics: to predict behaviour, act on predictions, and then feed those results back into the process in order to improve the predictions over time" (p. 5).

These descriptions beg fundamental questions. What data should be gathered for input to methods that generate learning analytics? Answering this question sets boundaries on and shapes, first, approaches to computations that underlie analytics and, second, what analytics can say about phenomena. For instance, ordinal (rank) data preclude using arithmetic operations on data, such as addition or division. If data are not ordinal, A cannot be described as greater than B, nor are transitive statements valid: if A > B and B > C, then A > C.

What properties of data bear on the validity of interventions based on learning analytics developed from the data? For example, determining that a learner's age, sex, or lab group predicts outcomes offers weak grounds for intervening without other data. None of these data classes are legitimately considered a direct, proximal (i.e., sufficient) cause of outcomes. Age and sex can't be manipulated; changing lab group may be impractical (e.g., due to scheduling conflicts with other courses or a job). And, notably, prediction does not supply valid grounds for inferring causality.

Who generates data? Who receives learning analytics? Learning ecologies involve multiple actors. Authors of texts and web pages vary cues they intend to guide learners about how to study; font styles and formats (bullet lists, sidebars that translate text to graphics)

1 https://tekri.athabascau.ca/analytics/

are examples. Instructional designers and instructors augment authors' works, for example, by setting goals for learning and elaborating content. They create and recommend schedules for learning; they control most opportunities for feedback to learners. Learners study solo and often form online cliques or study groups in which they exchange views about topics, share products of learning activities (e.g., questions, notes), and form and disengage from social units. The college or university strives to improve material and cyber infrastructure wherein other actors' work unfolds. Each category of actors generates data and is a legitimate candidate to receipt of learning analytics.

What are the temporal qualities – onset, duration, and offset – of collecting data, processing it, and delivering learning analytics? Will learners receive learning analytics as they work or will they need to be reminded of context when learning analytics are delayed? Are temporal delimiters positioned elastically or rigidly across a timeline of learning? Whose model of a learning episode – the analyst's or the learner's – matters?

Finally, what are learning analytics supposed to help improve? And, what standards should be used to gauge improvement? For example, if after receiving learning analytics a learner becomes more efficient in studying but achievement does not improve, is this a benefit? Is there value in freeing time for learners to engage in activities beyond academic assignments?

In this chapter, in keeping with a focus on *self*-regulated learning, the learner is positioned as the prime actor. Other actors' activities play roles as external conditions that may vary and, perhaps, be influenced by a learner's behaviour.

Self-Regulated Learning

A framework is useful to conceptualize learning analytics for self-regulated learning (SRL). When learners self-regulate their learning, they "actively research what they do to learn and how well their goals are achieved by variations in their approaches to learning" (Winne, 2010a, p. 472). One widely cited model elaborates features of SRL as four loosely sequenced recursive phases that unfold over the timeline of a task (Winne, 2011; Winne & Hadwin, 1998).

In phase 1, a learner surveys resources and constraints the learner predicts may affect how work on a learning task proceeds, the probability that specific actions bring about particular results, and the consequences of those activities. These factors can be located externally, in the learning environment or internal to the learner. Examples of external factors include access to information available from peers or in the Internet, software tools with functions designed to support learning in various ways, and time allowed for work on a task. Examples of internal factors include knowledge and misconceptions, interest in the task or topic, or a motivational disposition to interpret slow progress as a signal of low ability or of need to apply more effort (see Winne, 1995).

Having identified resources and constraints, in phase 2 a learner sets goals and plans how to approach them. Goals are standards a product should meet. Ipsative goals compare a learner's current results to earlier results; they measure personal growth (or decline). Criterion-referenced goals measure a product in relation to a fixed profile of task features or achievements in a domain. Norm-referenced goals position a learner's product relative to a peer's or a group's. Comparisons may be framed by a learner, an instructor, or other person. It is important to note that goals can target attributes of learning processes: which process is used, effort dedicated to carrying out a process, efficiency of a process, or increasing the probability a process yields a particular product. Goals also can be set in terms of products per se and their attributes; for example, number of pages written for an essay, anxiety reduced, or thoroughness of exposition. Plans describe actions a learner intends to carry out to approach goals. Every action potentially generates multiple products. Key products include information added to knowledge, errors corrected, gaps filled or misconceptions replaced. Products can also include the learner's perceptions about rate of progress, effort spent, opportunity to explore, or prospects to impress others.

In phase 3, the learner engages with the task by enacting planned operations. Working on a task inherently generates feedback that updates the task's conditions. Feedback may originate in the learner's external environment, such as when software beeps or a peer comments on a contribution to an online discussion. Or, feedback may arise internally as a result of the learner's monitoring work flow, such as when a search query is deemed unproductive because results were not what was expected or don't satisfy the need for particular information. Modest "course corrections" may result as the learner tracks updates to conditions across the timeline of a task. It is worth explicitly noting that goals can be updated.

Phase 4 is when the learner disengages from the task as such, monitors results in one or several of phases 1 to 3, and elects to make a large-scale change. Examples might be when a learner suspends work on solving a problem and returns to studying assigned readings with a goal to repair major gaps in knowledge; or, if re-studying is not predicted to be successful, the learner asks for help from the instructor. Changes may be immediately applied to the task, reshaping

Table 21.1. SMART Cognitive Operations

Operation	Description	Sample Traces
Search	Directing attention to particular information	Opening successive bookmarks. Using a search tool.
Monitor	Comparing information presentations in terms of standards	Highlighting text (the information highlighted meets a standard, e.g., important). Selecting a previously made note for review (e.g., judgment of learning).
Assemble	Relating items of information	Tagging. Assigning two bookmarks to a titled folder.
Rehearse	Maintaining or re-instating information in working memory	Reviewing a note. Copying, then pasting.
Translate	Transforming the representation of information	Paraphrasing. Describing a graph, equation, or diagram in words.

its multivariate profile in a major way. Or, plans for change may be filed for later use, what is called "forward reaching transfer."

A 5-slot schema describes elements within each phase of SRL. A first-letter acronym, COPES (Winne, 1997, summarizes the five elements in this schema. C refers to conditions. These are features the learner perceives influence work throughout phases of the task. For example, if there are no obvious standards for monitoring a product generated in phase 3, the learner may elect to search for standards or may abandon the task as too risky. Conditions fall into two main classes, as noted earlier. Internal conditions are the learner's store of knowledge about the topic being studied and about methods for learning, plus the learner's motivational and affective views about self, the topic, and effort in this context. External conditions are factors in the surrounding environment perceived to potentially influence internal conditions or two of the other facets of COPES, operations and standards.

O in the COPES schema represents operations. First-order or primitive cognitive operations transform information in ways that cannot be further decomposed. I proposed five such operations: searching, monitoring, assembling, rehearsing, and translating; the SMART operations (Winne, 2011). Table 21.1 describes each along with examples of traces – observable behaviour – that indicate an occurrence of the operation. Second- and higher-order descriptions of cognition, such as study tactics and learning strategies, are modelled as a pattern of SMART operations (see Winne, 2010a). An example study tactic is "Highlight every sentence containing a definition." An example learning strategy is "Survey headings in an assigned reading, pose a key question about each, then, after completing the entire reading assignment, go back to answer each question as a way to test understanding."

P is the slot in the COPES schema that represents products. Operations inevitably create products, though not always intended ones. A product can be uncomplicated, such as an ordered list of British monarchs, or complex, for example, an argument about privacy risks in social media or an explanation of catalysis. E represents evaluations of a product relative to standards, S, for products. Standards for a product constitute a goal.

Two further and significant characteristics of SRL are keys to considering how learning analytics can inform and benefit learners. First, SRL is an adjustment to conditions, operations, or standards. Thus, SRL can be observed only if data are available across time. Second, learners are agents. They regulate their learning within some inflexible and some malleable constraints, the conditions under which they work. As agents, however, learners always and intrinsically have choice as they learn. A learner may think, "I did it because I had to." A valid interpretation is that the learner elected to do it because the consequences forecast for not doing it were sufficiently unappealing as to outweigh whatever cost was levied by doing it.

The COPES model identifies classes of data with which learning analytics about SRL can be developed. In the next major section, I describe four main classes of data distinguished by their origin: traces, learner history, reports, and materials studied. In the following major section, I examine computations and reporting formats for learning analytics in relation to SRL. Together, these sections describe an architecture for learning analytics designed to support learners' SRL. In a final section, I raise several challenges to designing learning analytics that support SRL.

DATA FOR LEARNING ANALYTICS ABOUT LEARNING AND SRL

As learners work, they naturally generate ambient data (sometimes called accretion data; Webb, Campbell, Schwartz, & Sechrest, 1966). Ambient data arise in the natural course of activity. For example, clicking a hyperlink to open a web resource is data about a learner's cognition and motivation – based on whatever is the present context (perhaps the title of the resource), the learner forecast information in it has sufficient value to motivate viewing it. This click is a trace, a bit of ambient data that affords relatively strong inferences about one or more cognitive, affective, metacognitive, and motivational states and processes (CAMM processes; Azevedo, Moos, Johnson, & Chauncey, 2010). I offer two further examples of traces and inferences they afford. An explicit caution is the validity of inferences grounded in trace data should always be qualified by a probability <1.00 (certainty).

Highlighting a Sentence Fragment. To select particular text for highlighting among hundreds of sentences read in a typical study session, the learner metacognitively monitors attributes of information in the text relative to standards. Standards discriminate text to be highlighted from text that should not be highlighted. The learner might monitor information for "structural" features, such as definitions or principles; or for motivational/affective features, such as interestingness or novelty. Authors often attempt to signal information that should be highlighted using font styles (e.g., italics) or phrasing: "It is interesting that…" A highlight also traces that the learner plans to review highlighted text. Why else would the learner permanently mark selected text?

Reviewing a Note. Before reviewing a particular note, the learner engages in metacognitively monitoring what can be recalled about or what is understood about particular information. The learner chooses to review when recall is judged sufficiently inadequate perhaps because it is inaccurate, incomplete, or unclear. Searching for and re-viewing a particular note traces motivation to repair whatever problem the learner perceived. If the learner highlights information in the reviewed note, that identifies which particular information the learner had monitored and judged inadequate.

Four features describe ideal trace data gathered for generating learning analytics to support SRL. First, the sampling proportion of operations the learner performs while learning is large. Ideally, but not realistically, every operation throughout a learning episode is traced. Second, information operated on is identifiable. Third, traces are time stamped. Fourth, the product(s) of operations is (are) recorded. Data having this 4-tuple structure would permit an ideal playback machine to read data about a learning episode and output a perfectly mirrored rendition of every learning event and its result(s). With 4-tuple trace data, raw material is available to generate rich learning analytics.

In reality, every trace datum is at least mildly imperfect and slightly unreliable. For example, a highlight traces a monitoring operation and generates a product – the mark plus the content marked. At a future time, this marked content is a condition that facilitates focused review. What may not be clearly revealed by a highlight is the standard(s) the learner used to select that content. Better designed traces can reduce this gap. If learners are invited to tag content they highlight – interesting, important, unclear, project1, tellMike, – each tag the learner applies exposes a standard used to metacognitively monitor the information highlighted. In some cases, a tag reveals a stronger signal about a plan – e.g., use this content in project1, in the next chat tell Mike about this content.

Learner History

Instruments for recording traces to mirror the history of a learner's activities are available in at least three environments: paper systems, learning management systems, and systems that offer learners tools for studying "on the fly."

Paper Systems. In a paper-based learning environment, some examples of traces include content highlighted, notes, marginalia such as !, ?, and √ added to the whitespace of textbook pages, a pile of books or papers stacked in order of use (e.g., the topmost was most recently used), and post-it tabs of various colours attached to pages in a notebook.

Consider the ? symbol a learner may write in the margin of a textbook page. This symbol traces that the learner metacognitively monitored the meaning of content and judged it confusing or lacking information needed to fully grasp it. A further inference is available. Why would the learner spend effort to write the ? symbol in the margin? Content could be judged confusing or incomplete without recording a symbol. Odds are the learner is motivated to and plans to repair this gap, and return to context surrounding the text to improve understanding.

While tracing in a paper-based environment is easy for learners to do, it is hugely labour intensive to gather and prepare paper-bound trace data for input to methods that compute learning analytics. In software-supported environments, this burden is greatly eased.

Learning Management Systems. Today's learning management systems seamlessly record several time-stamped traces of learners' work, such as logging in and out, resources viewed or downloaded, assignments uploaded, quiz items attempted, and forum posts to anyone or to particular peers. By adding some simple interface features, goals can be inferred. For example, clicking a button labelled "practice test" traces a learner's judgment that knowledge is lacking or certitude is below a threshold of confidence. Aggregate trace data can support inferences about 1) learners' preferred work schedules that mildly support inferences about procrastination, 2) which resources are judged more relevant or appealing, 3) motivation to calibrate judgments of learning and efficacy, and 4) value attributed to contributing, acquiring, or clarifying by exchanging information with peers.

Traces gathered across the time stream can mark when learners first study a resource, if and when they review a resource, if and when they choose to self test, and when they take a test for marks. Coupled with other data about factors such as credit hours completed or the characteristics of peers with whom information is exchanged, traces like these provide raw material for building models about how learners self-regulate the study-review-practice-test cycle (Arnold & Pistilli, 2012; Delaney, Verkoeijen, & Spirgel, 2010; Dunlosky & Rawson, 2015).

When instructors or institutions require students to use a learning management system, ambient data are generated in the course of everyday use of the system. Costs incurred to collect trace data and prepare them for input to computations that generate learning analytics are slight.

Most learning management systems lack precision in traces with respect to tracking operations learners carry out as they study or review, and which particular information they study and review. A time-stamped trace that a resource was downloaded provides no information about whether the learner studied that content, not to mention how the learner studied it.

Software Tools for Studying. Winne and Baker (2013) nominated a triumvirate of motivation, metacognition and SRL as "raw material for engineering the bulk of an account about *why* and *how* learners develop knowledge, beliefs, attitudes and interests" (p. 1). They noted three challenges to research on improving learning outcomes by mining trace data about these factors: operationalizing indicators, gathering data that trace these constructs and filtering noise that obscures signals about the constructs (see also Roll & Winne, 2015a).

Operationalizing indicators – traces of COPES – calls for software developers to exercise imagination in designing interfaces that optimize opportunities for gathering trace data while supporting experimentation about learning and without enforcing new or perturbing a learner's preferred work habits. Table 21.2 presents illustrations of opportunities to gather trace data in a context where the learner uses software tools to:

- Search for information in a library containing assigned readings, supporting resources provided by an instructor, and artifacts the learner creates (e.g., terms, notes).
- Select content in an assigned reading to highlight it or tag it.
- Make a note structured by a schema that records the annotation in a web form with slots tailored to a schema – e.g., TERM NOTE: term, definition, example, see also …; or DEBATE NOTE: claim, evidence, warrant, counterclaim, my position.
- Organize items in a folder-like directory.

Phase 4, strategic revision of tactics and strategies for learning, is not included in Table 21.2; it is addressed in the later section on Learning Analytics for SRL.

As Winne and Baker (2013) noted, "Self-regulated learning (SRL) is a behavioural expression of metacognitively guided motivation" (p. 3). Consequently, every trace records a motivated choice about how to learn. Beyond representing features of the COPES model, traces reveal learners' beliefs about worthwhile effort that operationalizes choices among alternative goals.

The Learner's Reports

Paper-based questionnaires (surveys) and live oral reports are prevalent choices of methods for gathering data about learning events. Oral reports can be obtained through interviews outside the temporal boundaries of a studying session or during learning-on-the-fly as think aloud reports.

In both paper-based (or electronically presented) questionnaires and oral reports, learners are prompted to describe one or more features of COPES. The nature of the prompt is critical because it establishes several external conditions that a co-operative learner uses to set standards for deciding what to report. A thorough review is beyond the scope of this chapter; see Winne and Perry (2000) and Winne (2010b). In general, because questionnaire data are only weakly contextual (e.g., When you study, how often do you/how important is it for you to …?) and because all forms of self-report data suffer loss, distortion, and bias due to frailties of human memory, they may not reliably indicate how a learner goes about learning in any particular study episode or how learning varies (is self-regulated) as conditions vary. Self-report data are important, however, because they do reliably reflect beliefs learners hold about COPES. Beliefs shape what learners attend to about tasks, about themselves, and about standards they set for themselves.

Materials Studied

Materials learners work with are sources of data about conditions that may bear on how they engage in SRL. Texts can be described by various analytics including

Table 21.2. Illustrative Traces and Inferences about Phases of SRL

Phase of SRL	Trace	Inference
1) Survey resources and constraints	Search for "marking rubric" or "requirements" at the outset of a study episode.	An internal condition, namely, a learner's expectation that guidance is available about the requirements for a task.
	Open several documents, scan each for 15–30 s, close.	Refreshing information about previous work, if documents were previously studied; or scanning for particular but unknown information.
2) Plan and set goals	Start timer.	Plan to metacognitively monitor pace of work.
	Fill in fields of a "goal" note with slots: goal, milestones, indicators of success.	Assemble a plan in which goals are divided into sub-goals (milestones), set standards for metacognitively monitoring progress.
3) Engagement	Select and highlight content.	Metacognitive monitoring, unknown standards.
	Select and tag content.	Metacognitive monitoring; the standard used to monitor is revealed by the tag applied (e.g., confusing, good point).
	Select a bigram (e.g., greenhouse gas, slapstick comedy) and create a term.	Metacognitive monitoring content for technical terminology, assembling the term with a definition.
	Select content and annotate it using a "debate note" form, filling in slots: claim, evidence, warrant, counterclaim, my position.	Metacognitive monitoring with the standard to test whether content is an argument + assemble and rehearse information about the argument.
	Open a note created previously.	Metacognitive monitoring knowledge relative to a standard of completeness or accuracy, judge knowledge does not meet the standard.
	Put documents and various notes into a folder titled "Project Intro."	Metacognitively monitor uses of content; The standard is "useful for the introduction to a project"; assembling elements in a plan for future work.

readability[2] and cohesion (e.g., Coh-Metrix[3]). Content can be indexed for the extent to which learners have had opportunity to learn it plus characteristics of what a learner learned from previous exposures. Materials a learner studies also can be identified for the presence of rhetorical features such as examples and multichannel presentations of information, such as a quadratic expression described in words (semantic), an equation (symbolic), and a graph (visual) forms.

LEARNING ANALYTICS FOR SRL

Learning analytics to support SRL typically will have two elements: a calculation and a recommendation. The calculation – e.g., notation about presence, count, proportion, duration, probability – is based on traces of actions carried out during one or multiple study episodes (Roll & Winne, 2015a). A numeric report may be conveyed along with or as a visualization. Examples might be a stacked bar chart showing relative proportions of highlights, tags and notes created while studying each of several web pages, a timeline marked with dots that show when particular traces were generated, and a node-link graph depicting relations among terms in a glossary (link nodes when one term is defined using another term) with heat map decorations showing how often each term was operated on while studying. This element directly or by transformation mirrors information describing COPES traced in the history of a learner's engagement. Table 21.3 presents illustrative trace data that might be mirrored.

A "simple" history of trace data mirrored back to a learner may be conditioned or contextualized by other data: features of materials such as length or a readability index, demographic data describing the learner (e.g., prior achievement, hours of extracurricular work, postal code), or other characterizations of learners such as disposition to procrastinate, degree in a social network (the number of people with whom this learner has exchanged information) or context for study (MOOC vs. face-to-face course delivery, opportunity to submit drafts for review by peers before handing in a final copy to be marked).

The second element of a learning analytic about SRL is a recommendation – what should change about how learning is carried out plus guidance about how to go about changing it. Learners can directly control three facets of COPES: operations, standards, and some conditions (Winne, 2014). Products are controllable only indirectly because their characteristics are function of 1) conditions a learner is able to and chooses to vary, particularly information selected for operations; and, 2) which operation(s) the learner chooses to apply in manipulating information. Evaluations are determined by the match of product attributes and the particular standards a learner adopts for those products. Recommendations about changing conditions, operations, or standards may be grounded in findings from data mining not guided by theory, by findings from research

2 See, for example, http://www.wordscount.info/readability.html
3 http://cohmetrix.com/

Table 21.3. Analytics Describing COPES Facets in SRL

COPES	Description
Conditions	Presence/absence of a condition within a learning episode Onset/offset along the timeline in a study episode or across a series of episodes
Operations	Frequency of SMART operations (see Table 21.1) Sequence, pattern, conditional probability one SMART operation relative to others
Product	Presence Completeness (e.g., number of fields with text entered in a note's schema) Quality
Standard	Presence Precision Appropriateness
Evaluation	Presence Validity

in learning science, nor a by combination.

Whether a recommendation is offered or not, change in the learner's behaviour traces the learner's evaluation that 1) previous approaches to learning were not sufficiently effective or satisfactory and 2) the learner predicts benefit by adopting the recommendation or an adaptation of it. In this sense, learning analytics update prior external conditions and afford new internal conditions. Together, a potential for action is created, but this is only a potential for two reasons. First, learners may not know how or have skill to enact a recommendation. Second, because learners are agents, *they* control their learning. As Winne and Baker (2013) noted:

> What marks SRL from other forms of regulation and complex information processing is that the goal a learner seeks has two integrally linked facets. One facet is to optimize achievement. The second facet is to optimize how achievement is constructed. This involves navigating paths through a space with dimensions that range over processes of learning and choices about types of information on which those processes operate. (p. 3)

Thus, learning analytics afford opportunities for learners to exercise SRL but the learner decides what to do. There is an important corollary to this logic. If a learning analytic is presented without a recommendation for action, an opportunity arises for investigating options a learner was previously able to exercise on his or her own and, now, chooses to exercise. In other words, motivation and existing tactics for learning can be assessed by analytics that omit recommendations and guidance for action.

CHALLENGES FACING LEARNING ANALYTICS ABOUT SRL

Research on learning analytics as support for SRL is nascent. The field has just begun to map frontiers, including what to trace, instrumentation for gathering traces, interfaces that optimize gathering data without overly perturbing learning activities, computational tools for constructing analytics about SRL that meld trace data with other data, scheduling delivery of learning analytics, and features of information conveyed in learning analytics (Baker & Winne, 2013; Roll & Winne, 2015b). Amidst these many topics, several merit focused exploration.

Grain Size

Features of learning events can be tracked at multiple grain sizes ranging from individual keystrokes and clicks executed along a timeline marked off in very fine time units (e.g., tens of milliseconds) to quite coarse grain sizes (e.g., the URL of a web page and when it loads, the learner's overall score on a multi-item practice quiz). Different methods for aggregating fine-grained data will represent features of COPES differently. While this affords multiple views of how learners engage in SRL, several questions arise.

First, how will depictions of SRL and recommendations for adapting learning vary across learning analytics formed from data at different grain sizes? An analogy might be made to chemistry. Chemical properties and models of chemical interactions vary depending on whether the unit is an element, a compound, or a mixture. Consider two grain sizes for information that is manipulated with an assembling operation: 1) snippets of text selected for tagging when studying a web page, and 2) entire artifacts – quotes, notes, and bookmarks – that a learner files in a titled (tagged) folder. Future research may reveal that assembling at one grain size has different implications for learning relative to assembling at another grain size.

If grain size matters, one implication is that approaches to forming learning analytics may benefit by considering not only whether and which operations are applied – what a learner does – but also characteristics of information to which operations are applied.

Learning analytics for SRL may benefit by blending counts and other quantitative descriptions of COPES with semantic, syntactic, and rhetorical features of conditions, products, and standards.

Because coarser grained reflections of SRL generally, but not necessarily, are built up using finer grained data, another issue arises in developing and using statistical calculations. Statistical descriptions that describe relationships among larger-grained features of learning and SRL, such as correlation and distance metrics, may share finer-grained constituents. This inherently introduces part-whole relationships. Will that matter?

Time

Excepting research in learning science that investigates how achievement covaries with time spans between episodes of studying, reviewing, and taking tests (Delaney et al., 2010), the phenomenon of forgetting (Murayama, Miyatsu, Buchli, & Storm, 2014) and loss of knowledge across the summer vacation (Cooper, Nye, Charlton, Lindsay, & Greathouse, 1996), time data has been underused. Traces and other data available to learning analytics commonly can be supplemented with time stamps. Much research remains to investigate how temporal features of COPES and coarser-grained descriptors may play useful roles in learning analytics about SRL as a process that unfolds within each studying episode and across a series of episodes. One focus for this research is identifying patterns in COPES events across time (Winne, Gupta, & Nesbit, 1994). Vexing questions here are how to define the span of a time window within which patterns are sought and the degree to which non-focal events intervening in an encompassing pattern can be identified and filtered out (see Zhou, Xu, Nesbit, & Winne, 2011). Another key topic relating to time is investigating when learning analytics should be delivered: in real time (i.e., approximately instantaneously following an event or identification of a pattern), on demand (by learners or instructors), or at punctuated intervals (e.g., weekly)?

Generalization

Learning science strives to balance the accuracy of descriptions about particular learning events in contrast to describing how learning events relate to outcomes, which requires ignoring details to allow generalizing over specific events. When data and time stamps at very fine-grain sizes are available about the course of studying over time, accuracy of description is maximized. How should generalizations be formed, tested, and validly interpreted as accuracy is deliberately compromised (see Winne, 2017)?

The goal of education is development – of knowledge, interest, confidence, critical thinking, and so on. If education succeeds, each learner changes over time, and changes quite likely vary among peers. Even if there is genuinely big data, at very fine grain sizes of data, it is statistically very unlikely any two learners' data signatures perfectly match. Learning analytics face a challenge to find balance between accuracy and generalization when describing one learner's ipsative development or the match of that learner's "learning signature" to others'. The field of learning analytics will benefit from frequent consideration of this challenge.

ACKNOWLEDGEMENTS

This work was supported by grants to Philip H. Winne from the Canada Research Chairs Program, and the Social Sciences and Humanities Research Council of Canada 410-2011-0727 and 435 2016-0379.

REFERENCES

Arnold, K. E., & Pistilli, M. D. (2012). Course Signals at Purdue. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 267–270). New York: ACM. doi:10.1145/2330601.2330666

Azevedo, R., Moos, D. C., Johnson, A. M., & Chauncey, A. D. (2010). Measuring cognitive and metacognitive regulatory processes during hypermedia learning: Issues and challenges. *Educational Psychologist*, 45(4), 210–223.

Baker, R. S. J. d., & Winne, P. H. (Eds.). (2013). Educational data mining on motivation, metacognition, and self-regulated learning [Special issue]. *Journal of Educational Data Mining*, 5(1).

Cooper, H., Nye, B., Charlton, K., Lindsay, J., & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66(3), 227–268.

Delaney, P. F., Verkoeijen, P. P., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of Learning and Motivation*, 53, 63–147.

Dunlosky, J., & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: Tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology*, 1(1), 72–78.

Educause (n.d.) Next generation learning initiative. http://nextgenlearning.org

Eckerson, W. W. (2006). *Performance dashboards: Measuring, monitoring, and managing your business*. Hoboken, NJ: John Wiley & Sons.

Elias, T. (2011, January). *Learning analytics: Definitions, processes and potential*. http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Murayama, K., Miyatsu, T., Buchli, D., & Storm, B. C. (2014). Forgetting as a consequence of retrieval: A meta-analytic review of retrieval-induced forgetting. *Psychological Bulletin*, 140(5), 1383–1409.

Roll, I., & Winne, P. H. (2015a). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, 2(1), 7–12.

Roll, I., & Winne, P. H. (Eds.). (2015b). Self-regulated learning and learning analytics [Special issue]. *Journal of Learning Analytics*, 2(1).

Siemens, G. (2010, August 25). *What are learning analytics?* http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures*. Skokie, IL: Rand-McNally.

Winne, P. H. (1995). Inherent details in self-regulated learning. *Educational Psychologist*, 30, 173–187.

Winne, P. H. (1997). Experimenting to bootstrap self-regulated learning. *Journal of Educational Psychology*, 89, 397–410.

Winne, P. H. (2010a). Bootstrapping learner's self-regulated learning. *Psychological Test and Assessment Modeling*, 52, 472–490.

Winne, P. H. (2010b). Improving measurements of self-regulated learning. *Educational Psychologist*, 45, 267–276.

Winne, P. H. (2011). A cognitive and metacognitive analysis of self-regulated learning. In B. J. Zimmerman and D. H. Schunk (Eds.), *Handbook of self-regulation of learning and performance* (pp. 15–32). New York: Routledge.

Winne, P. H. (2014). Issues in researching self-regulated learning as patterns of events. *Metacognition Learning*, 9(2), 229–237.

Winne, P. H. (2017). Leveraging big data to help each learner upgrade learning and accelerate learning science. *Teachers College Record*, 119(3). http://www.tcrecord.org/Content.asp?ContentId=21769.

Winne, P. H. (in press). Cognition and metacognition in self-regulated learning. In D. Schunk & J. Greene (Eds.), *Handbook of self-regulation of learning and performance*, 2nd ed. New York: Routledge.

Winne, P. H, & Baker, R. S. J. d. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, 5(1), 1–8.

Winne, P. H., Gupta, L., & Nesbit, J. C. (1994). Exploring individual differences in studying strategies using graph theoretic statistics. *Alberta Journal of Educational Research*, 40, 177–193.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum Associates.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 531–566). Orlando, FL: Academic Press.

Zhou, M., Xu, Y., Nesbit, J. C., & Winne, P. H. (2011). Sequential pattern analysis of learning logs: Methodology and applications. In C. Romero, S. Ventura, S. R. Viola, M. Pechenizkiy & R. Baker (Eds.), *Handbook of educational data mining* (pp. 107–121). Boca Raton, FL: CRC Press.

Chapter 22: Analytics of Learner Video Use

Negin Mirriahi[1], Lorenzo Vigentini[2]

[1]School of Education & Teaching Innovation Unit, University of South Australia, Australia
[2]School of Education & PVC (Education Portfolio), University of New South Wales, Australia
DOI: 10.18608/hla17.022

ABSTRACT

Videos are becoming a core component of many pedagogical approaches, particularly with the rise in interest in blended learning, flipped classrooms, and massive open and online courses (MOOCs). Although there are a variety of types of videos used for educational purposes, lecture videos are the most widely adopted. Furthermore, with recent advances in video streaming technologies, learners' digital footprints when accessing videos can be mined and analyzed to better understand how they learn and engage with them. The collection, measurement, and analysis of such data for the purposes of understanding how learners use videos can be referred to as *video analytics*. Coupled with more traditional data collection methods, such as interviews or surveys, and performance data to obtain a holistic view of how and why learners engage and learn with videos, video analytics can help inform course design and teaching practice. In this chapter, we provide an overview of videos integrated in the curriculum including an introduction to multimedia learning and discuss data mining approaches for investigating learner use, engagement with, and learning with videos, and provide suggestions for future directions.

Keywords: Video, analytics, learning, instruction, multimedia

With the rise in online and blended learning, massive open and online courses (MOOCs) and flipped classroom approaches, the use of video has seen a steady increase. Although much research has been done, particularly focusing on psychological aspects, the educational value, and the user experience, the advancements of the technology and the emergence of analytics provide an opportunity to explore and integrate not only how videos are used in the curriculum but whether their adoption has contributed towards learner engagement or learning (Giannakos, Chorianopoulos, & Chrisochoides, 2014). Educators are choosing to bring videos into their courses in a variety of ways to meet their particular intentions. This is occurring not only in higher education, continuing professional development, and the K–12 sectors but also in corporate and government training (Ritzhaupt, Pastore, & Davis, 2015). Therefore, it is important to evaluate or investigate how learners are using and engaging with videos in order to inform future modifications or advances in how they are integrated into the curriculum.

The use of videos in the curriculum stems from earlier use of multimedia in learning environments dating back several decades. Hence, before exploring how videos are integrated into the curriculum or identifying methods to investigate how learners use them, it is important to consider prior research conducted on multimedia learning. This chapter begins with a discussion of related work, specifically multimedia learning and strategies for evaluating learning with multimedia. This is followed by methodological considerations including video types, the ways videos can be integrated into the curriculum, and the data mining approaches that can be applied to understand use, engagement, and learning with videos. The final section summarizes the chapter and offers directions for further exploration.

RELATED WORK

What Do We Know about Learning with Multimedia and Interactive Courseware?

"People learn better from words and pictures than from words alone" is the key statement driving the popular work of Mayer (2009) on multimedia learning. Videos are a form of multimedia and therefore this chapter will leverage a wealth of research exploring their effectiveness for learning. Since the introduction of computers and instructional technology in education, both the research on and the development of interactive course materials, followed the trends and shifts of beliefs in psychological and educational research and can be identified within one of the following three phases/perspectives:

1. **Behaviourist:** presenting an objectivist view of knowledge and instructional design features focusing on serial structuring of material, program/delivery control, and regular review and testing against specified criteria – from Skinner's (1950) radical behaviourism to Gagne's (1965) tenets on the conditions of learning.
2. **Cognitive:** focusing on the factors affecting effective learning and teaching with attention to information processing and the characteristics of the learner, the teachers, and the learning environment (Keller, 1967; McKeachie, 1974).
3. **Constructivist:** knowledge-building with a focus on the interdependence of social and individual processes in the co-construction of knowledge (Palincsar, 1998).

In between these often entrenched perspectives, the key issue has been the definition of how much instruction affects learning (Lee & Anderson, 2013) and, in particular, how much the instructivist and constructivist approaches deriving from the three perspectives facilitate active learning. According to the behavioural perspective, learning can be efficiently accomplished with a strong set of instructions and a specific sequence of learning (Kirschner, Sweller, & Clark, 2006; Lee & Anderson, 2013) but there may be a trade-off between efficiency and effectiveness (Atkins, 1993). Instead, the key weakness of the cognitive orientation is the articulation or provision of suitable metacognitive frameworks to support learning. Constructivist and connectivist models of learning are student-centred in nature and imply a level of self-directedness and self-regulation in order to navigate through the teaching material to determine the most suitable learning pathway. Notwithstanding the philosophical perspective taken, "what seems to be missing are models of learning appropriate for the design opportunities offered by new technologies" (Atkins, 1993, p. 252) and this includes videos and multimedia.

Practitioners and instructional designers find comfort in Gagne's (1965) model of instructional events and his classification of types of learning outcomes because of their relative ease of adoption and use (Reeves, 1986). Furthermore, the concept of mastery learning (Bloom, 1968) has attracted a large amount of research supporting its effectiveness (Guskey & Good, 2009; Kulik, Kulik, & Bangert-Drowns, 1990) and together with the five main elements of Keller's (1967) Personalized System of Instruction (PSI), as noted in Figure 22.1, strongly influenced instructional design and learning sciences.

Mayer (2009) attempts to summarize the wealth of knowledge on multimedia learning accrued over the past four decades in the formulation of 12 principles, as noted in Figure 22.2. These principles provide insight into the way people learn with multimedia, grounded in evidence from psychology, instructional design, and the learning sciences. Being aware of the positive and negative design features and their known effects on learning is very important when an instructor is integrating videos into their teaching.

Another aspect to consider, described in more detail below in Data-Mining Approaches to Videos, is the issue of engagement and how learning relates to patterns of engagement. Although Mayer and colleagues demonstrated the effects of certain features of the medium on learning, when moving from a lab context to real life, the extent to which a learner interacts with the medium is an important aspect. This is mediated not only by the characteristics of the medium, but also by the individual preferences and approaches to learning, which make it quite hard to clearly disentangle the relation between the volume or amount of engagement with videos (i.e., the interaction) and the learning, which is tied to the mode of assessing learning.

- The go-at-your-own-pace feature, which permits a learner to move through the course at a speed commensurate with his ability and other demands upon his time
- The unit-perfection requirement for advancement, which lets the learner advance to new material only after demonstrating mastery of that which preceded
- The use of lectures and demonstrations as vehicles of motivation, rather than sources of critical information
- The related stress upon the written word in teacher–learner communication
- The use of proctors, which permits repeated testing, immediate scoring, almost unavoidable tutoring, and a marked enhancement of the personal-social aspect of the educational process

Figure 22.1. Keller's (1967) elements of the PSI (Personalized System of Instruction).

1. **Coherence Principle** – People learn better when extraneous words, pictures, and sounds are excluded rather than included.
2. **Signalling Principle** – People learn better when cues that highlight the organization of the essential material are added.
3. **Redundancy Principle** – People learn better from graphics and narration than from graphics, narration, and on-screen text
4. **Spatial Contiguity Principle** – People learn better when corresponding words and pictures are presented near rather than far from each other on the page or screen.
5. **Temporal Contiguity Principle** – People learn better when corresponding words and pictures are presented simultaneously rather than successively.
6. **Segmenting Principle** – People learn better from a multimedia lesson, which is presented in user-paced segments rather than as a continuous unit.
7. **Pre-training Principle** – People learn better from a multimedia lesson when they know the names and characteristics of the main concepts.
8. **Modality Principle** – People learn better from graphics and narrations than from animation and on-screen text.
9. **Multimedia Principle** – People learn better from words and pictures than from words alone.
10. **Personalization Principle** – People learn better from multimedia lessons when words are in conversational style rather than formal style.
11. **Voice Principle** – People learn better when the narration in multimedia lessons is spoken in a friendly human voice rather than a machine voice.
12. **Image Principle** – People do not necessarily learn better from a multimedia lesson when the speaker's image is added to the screen.

Figure 22.2. Mayer's (2009) multimedia design principles.

Key Considerations in the Multimedia Literature

A recent review of the literature on video-based learning between 2003–2013 (Yousef, Chatti, & Schroeder, 2014) provided a useful overview of the types of studies conducted. This categorization is reproduced in Figure 22.3 below.

This provides a good starting point to make sense of the most recent research directions, but mostly ignores the research produced in the previous five decades, culminating in Mayer (2009), who identifies six major strands relevant to multimedia and video in learning and education: 1) perception and attention; 2) working memory and memory capacity; 3) cognitive load theory; 4) knowledge representation and integration; 5) learning and instruction (including learning styles, approaches, and instructional methods); and 6) self-regulation of learning. These areas provide the theoretical backdrop necessary to understand, identify, and select adequate analytics (intended here as both methods and metrics) to demonstrate the effectiveness of videos for learning. In particular, work done on the first three areas provides essential parameters to determine the way in which a learner may *interact* and *engage* with videos, and the second set of three provides useful data to understand the way in which *learning* from and with videos occurs, all illustrated in Figure 22.4.

Relating back to the taxonomy of Yousef and colleagues (2014), the notion of *effectiveness* fits in the broader learning space (Figure 22.4) and is at the centre of the discussion and the evaluation of how effectiveness can

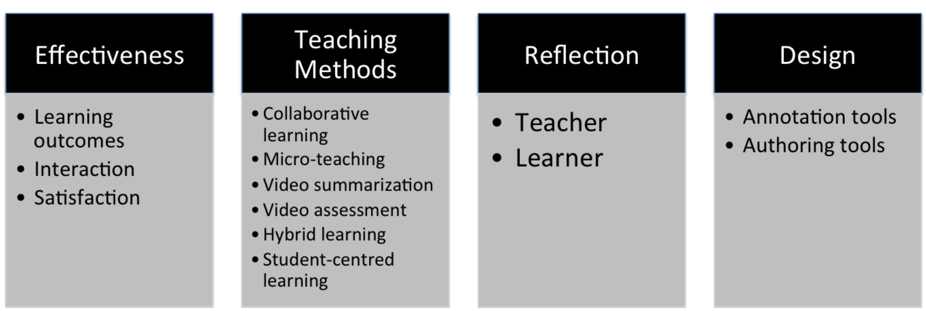


Figure 22.3. Overview of the video-based literature 2003–2013 (adapted from Yousef et al., 2014).

be applied to instruction, the learners, and the tools (specifically videos). There is a direct connection between instruction and the learner: this is exemplified in what an instructor does to facilitate learning to respond to the learners, represented under teaching methods as noted in Figure 22.3 (Yousef et al., 2014). Furthermore this is extended in Figure 22.4 to illustrate the relationship between both instruction and the learner and instruction and the learning tools – i.e., the video. In the interaction between instruction and the learner, elements such as learning styles, approaches, and instructional methods alongside self-regulated learning affect the effectiveness of instruction and learning. Instruction is affected by the instructional methods and cognitive load, and multimedia learning theory. The direct relation between the instruction and the instructional tools such as the resources, activities, supporting and evaluation tools – in this particular case the use of videos – is partly present in Yousef and colleagues' (2014) taxonomy under "design," but the important reference to the learner is missing, especially when students not only consume videos, but also produce them (Juhlin, Zoric, Engström, & Reponen, 2014). Finally, the dual relationship between the learning and the video is affected by learners' perception, attention, their working memory and capacity as well as their preferences driven by the affordances of the videos. Figure 22.4 also shows some key metrics that could be applied to investigate the relationships between instruction, the learner, and the video when measuring effectiveness.

Evaluation Methods to Investigate the Effectiveness of Videos on Learning

In order to evaluate the effectiveness of multimedia and videos, together with experimental (e.g., lab) studies, a plethora of published research proposes a *comparative approach* (see Data-Mining Approaches to Videos below), or a "horserace" model for evaluating the comparison of a mythical "traditional instruction" with the latest innovations in instructional technology tools (Reeves, 1986, 1991). Although experimental studies have a certain appeal and credibility, research studies adopting experimental or quasi-experimental designs comparing instructional technologies have produced very few useful outcomes. Literature reviews and meta-analyses have recognized this phenomenon as the "no significant differences" problem (Joy & Garcia, 2000; Oblinger & Hawkins, 2006; Russell, 1999). Videos have been subjected to similar comparative studies since the 1980s – initially with a focus on videodiscs and interactive videos, and later with computer-based instruction, video animations, documentaries, and video-recorded presentations or lectures. The debate on the influence of media on learning has been well represented by the opposing views of Clark (1983, 1994) and (Kozma, 1991, 1994). Clark (1994) argues that media does not influence learning under any condition; however, "learning is caused by the instructional methods embedded in the media presentation" (p. 26). Notably, instructional methods were defined as "any way to shape information that activates, supplants, or compensates for the cognitive processes necessary for achievement or motivation"

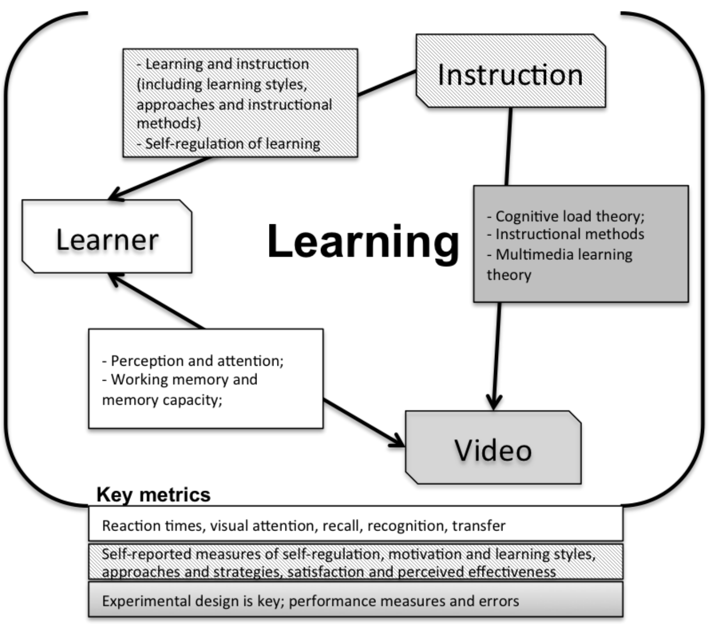


Figure 22.4. Interconnections between the learner, instruction, and video with reference to some of the key metrics used in the literature.

(p. 23). On the other side of the debate, Kozma (1991, 1994) argues that media and methods are intertwined and dependent on each other:

> From an interactionist perspective, learning with media can be thought of as a complementary process within which representations are constructed and procedures performed, sometimes by the learner and sometimes by the medium [...] media must be designed to give us powerful new methods, and our methods must take appropriate advantage of media's capabilities (1994, pp. 11, 16).

Within this "Great media debate," Tennyson (1994) argues, "a scientist is never satisfied with the current state of affairs, but is always and foremost challenged by extending knowledge" (p. 15). He asserts that a scientist turns into an advocate when statistically significant results are found and the newly found approach is adopted to tackle the world's complexity: this is termed the "big wrench" approach. "The advocate, with the big wrench in hand, sets out to solve, suddenly, a relatively restricted number of problems. That is, all of the formerly many diverse problems, now seem to be soluble with the new big wrench (or panacea)" (p. 16). This should provide a stark warning against the temptation of focusing too much or exclusively on one method of evaluation, (e.g., analytics) as the potential "big wrench" used to make sense of learning with and from videos in education. Instead, a range of approaches should be used to investigate and evaluate use, engagement, and learning with videos. Such strategies will be explored in Data-Mining Approaches to Videos, below.

METHODOLOGICAL CONSIDERATIONS

Video Types

Videos have become increasingly important to provide varied pedagogical opportunities to engage learners and respond to the growing need for flexible, blended, and online learning modes. There are two broad categories of video use: synchronous and asynchronous. The former provides a real-time opportunity for learners and instructors to engage with one another simultaneously through virtual classrooms, live webcasts, or video feeds. The latter supports self-paced learning and is primarily an individual interaction between the medium and the learner. Asynchronous videos are becoming more common and vary from the capture of an in-class lecture, to the recording of an educator's talking head or their audio of a lecture accompanied by slides or images illustrating core concepts (Owston, Lupshenyuk, & Wideman, 2011). Such lecture videos can be a variety of durations and have become more mainstream with the introduction of automatic lecture recordings in many lecture halls facilitated by technologies such as Echo360[1], Opencast[2], and Kaltura[3] minimizing the resources and time required of the educator to produce the videos. While in some learning contexts, these videos are provided to learners as supplemental resources, many educators are adopting flipped classroom approaches whereby information-transmission is done through required video lectures prior to class time, providing time in class for collaborative and active learning activities. Further, lecture videos have also gained momentum with their recent availability through streaming platforms such as YouTube, Apple's iTunes U program, and the Khan Academy where a vast variety of videos covering various disciplines and concepts are available. MOOCs have also contributed to the widespread adoption of lecture videos. For many MOOC providers (e.g., Coursera, Udacity, EdX), a core functionality is the provision of video streaming, providing much of the course content via videos supported by quizzes, forums, and readings (Diwanji, Simon, Marki, Korkut, & Dornberger, 2014; Li, Kidzinski, Jermann, & Dillenbourg, 2015). These particular MOOCs, which focus heavily on video lectures and individual mastery of content (e.g., via quizzes with immediate feedback), follow a cognitive-behaviourist approach, often referred to as xMOOCs (Conole, 2013), and have largely developed since 2012 (Margaryan, Bianco, & Littlejohn, 2015). While in many cases higher education videos are the result of live recording with what is available (i.e., automatic lecture recording or desktop recording using screen capture and audio recording), in MOOCs, videos tend to be scripted, recorded, and edited with high-end equipment and slick production values (Guo, Kim, & Rubin, 2014; Ilioudi, Giannakos, & Chorianopoulos, 2013; Kolowich, 2013).

Lecture videos are not the only type of educational videos being integrated in the curriculum. Video recordings of learners' own performances or engagement with an activity are used for self-reflection, peer and instructor feedback, and goal-setting purposes. For example, pre-service teachers have viewed recordings of their own teaching scenarios and made notes or markers on particular segments for their own self-reflection purposes or to provide feedback to peers facilitated by video annotation software such as the Media Annotation Tool (MAT) (Colasante, 2010, 2011). The use of video recordings for learner reflection and critical analysis has also been used in medical education whereby learners view recordings of their consultations with simulated patients and explain their behaviour and note areas of improvement linked

[1] Echo360 Active Learning Platform http://echo360.com
[2] Opencast http://www.opencast.org
[3] Kaltura http://corp.kaltura.com

to specific time-codes in the video using annotation software, DiViDU (Hulsman, Harmsen, & Fabriek, 2009). In the performing arts discipline, videos of learners' own performances have been used for self-reflection purposes (Daniel, 2001) and more recently coupled with video annotation software, CLAS, for learners to make time-stamped and general comments related to their performance (Gašević, Mirriahi, & Dawson, 2014; Mirriahi, Liaqat, Dawson, & Gašević, 2016; Risko, Foulsham, Dawson, & Kingstone, 2013).

Video in the Curriculum

Although videos are included in the curriculum in various ways, it is not often transparent whether the integration of videos into the course has been effective or requires further refinement. Whether as supplemental resources, core components of flipped classroom approaches or MOOCs, or used for reflective practice and peer feedback, it is important to understand how learners engage with the videos and how it contributes to their learning experience (Giannakos et al., 2014). To date, numerous studies have been conducted in various educational settings exploring the effectiveness of videos in the curriculum (Giannakos, 2013; Yousef et al., 2014). However many of the earlier studies have largely relied on learners' and educators' self-reports rather than objective data. Relying solely on self-reports can lead to potential inaccurate recall of learners' prior behaviour (Winne & Jamieson-Noel, 2002) or lead to social-desirability bias whereby learners provide the expected response rather than the most accurate one (Beretvas, Meyers, & Leite, 2002; Gonyea, 2005). Recent advances in learning analytics and data mining techniques, however, can provide more objective and authentic data regarding learners' actual use of learning technologies by analyzing their digital footprints (Greller & Drachsler, 2012). Hence, mining the data from learners' use of videos as a complement to other data sources (e.g., assessment scores, surveys, etc.) (Giannakos, Chorianopoulos, & Chrisochoides, 2015) can help begin to uncover how learners actually use videos and how they contribute to their learning experience. Leveraging the trace or clickstream data available from learners' use of videos, which can be termed *video analytics*, has become more readily available in recent years from streaming video platforms (YouTube, Vimeo) or MOOC providers (Udacity, Coursera, FuturLearn, EdX).

Broadly, we define *video analytics* as the collection, measurement, and analysis of data from learners' use of videos for the purposes of understanding how they engage with them in learning contexts. This provides the opportunity to mine learners' actual use of the videos alongside data collected from other online activities such as quizzes or annotations to explore when and how learners engage with the videos and their associated activities. Aggregating such data with performance measures (e.g., grades or scores) can help identify any impact on learning outcomes while collecting information about learners' intentions or motivations through self-reports can help explain learner use. Collectively, these varied data sets can help reveal whether learners are engaging or using video technologies as intended by the course design or if further revisions to the pedagogical approach are required to better meet the intended outcomes of the learning and teaching strategy (Pardo et al., 2015).

In the next section, we discuss various approaches to studying learner use of video technologies (using video analytics alongside other data collection methods) to begin to understand patterns in learning and engagement.

DATA MINING APPROACHES TO VIDEOS

Data mining is commonly defined as the process of collecting, searching through, and analyzing a large amount of computerized data, in order to discover patterns, trends, or relationships (Witten & Frank, 2005; Romero & Ventura, 2010; Peña-Ayala, 2014). This is done through a combination of tools and methods used in statistics and artificial intelligence (AI). The algorithms driving the *mining* process derive from a field of research in AI termed knowledge discovery and machine learning: the broad categories of machine learning and the algorithms associated with these are represented in Figure 22.5.

Data mining has long been used to study multimedia and video. This is partly because of the relative ease in creating new content and the availability of web-based video streaming services to distribute videos. When one looks at applications and mining techniques for videos, there are two major strands of work: 1) making sense of the content of the video and 2) exploring how learners use videos. In the next two sections, we will explore some of these methods and techniques.

Figure 22.5 provides an overview of the connections of machine learning algorithms and applications applicable to *video analytics*. It provides a distinction between supervised and unsupervised machine learning, leading to the specific types of algorithms used to analyze the data from different applications of video (e.g., learners' interaction with video content and their use of videos). Depending on the nature of the data sources (for example, usage and in-video behaviours, or frame analysis), different families of algorithms are more appropriate for making sense of the data.

In this chapter, we will not dwell on the effectiveness of algorithms, but briefly describe some applications

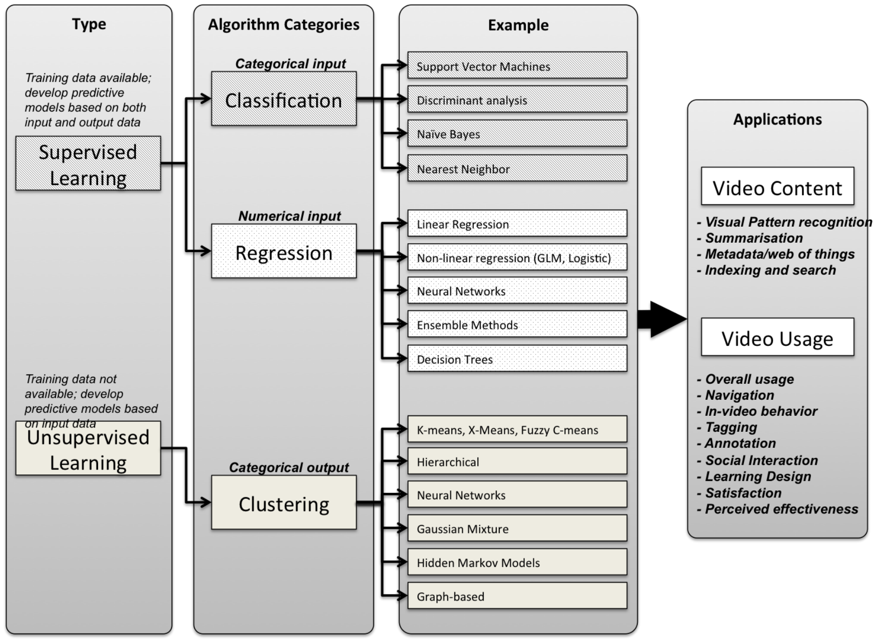


Figure 22.5. Interconnections between the learner, instruction, and video with reference to some of the key metrics used in the literature.

of video content mining, giving particular attention to usage. Table 22.1 provides an explicit mapping of existing literature, algorithms, types of interactions, and features used in the analysis.

Mining Applied to Content

Making sense of videos is a complex problem that leverages advances in automated content-based methodologies such as visual pattern recognition (Antani, Kasturi, & Jain, 2002), machine learning (Brunelli, Mich, & Modena, 1999), and human-driven action (Avlonitis, Karydis, & Sioutas, 2015; Chorianopoulos, 2012; Risko et al., 2013). The latter can be individual use of video resources or the social metadata (tagging, sharing, and social engagement). Video indexing, commonly used to make sense of video content, is based on three main steps: 1) video parsing, 2) abstraction, and 3) content analysis (Fegade & Dalal, 2014). Furthermore, given the exponential growth of video content, the problems of navigation (or searching within content) and summarization can also be resolved using content analytics (Grigoras, Charvillat, & Douze, 2002; He, Grudin, & Gupta, 2000).

The relevance of this work can be seen in the ability to characterize and present video content to learners and the methods to integrate this medium with teaching and instruction. For example, a better way of guiding learners to key points in a video or providing learners with ways to regulate their own learning with videos is to provide a navigational index much like a table of contents or glossary to allow learners to jump to the most relevant part of video. One way of providing this is using video annotation software, which provides instructors and learners with the option of flagging particular time-stamped parts of a video for later review and to gauge their learning in relation to others by viewing other's annotations or flags (Dawson, Macfadyen, Evan, Foulsham, & Kingstone, 2012). Given that new users of websites and applications tend to watch videos and skip text while more expert users skip videos and scan the associated text (Johnson, 2011), this creates interesting design problems and questions: by aggregating learner engagement or use of videos, could the "crowd-sourced" expertise of learners provide automated or user-driven instructional support scaffolding novice or less experienced learners or is the adoption of machine learning more effective? Although there is much evidence in favour of machine driven methods (i.e., EDM and AIED communities), the problem of knowledge representation and transfer remains a crucial one.

Another source of accessible metadata about videos came about with assistive technologies and the synching of text transcripts related to videos. For example, Ed-X displays both video and transcripts on the same page, whilst YouTube has recently introduced an automatic caption tool to create subtitles.

Mining Applied to Usage: Logs of Activity to Measure Interaction

The extraction of trace or log data from learners' use of video technologies and the analysis to understand learning processes or engagement is still at an early stage both in terms of a research discipline (e.g., learning analytics and educational data mining) but also in terms of how it is used to inform teaching practice. As noted by Giannakos, Jaccheri, and Krogstie (2015), further experimentation with methodological approaches is needed to advance the area. Yet, there are a growing number of studies exploring how learners learn and engage with video technologies using educational data mining and learning analytics methods alongside more traditional data collection approaches (e.g., questionnaires, observations, and interviews). To provide an overview we specifically looked at published literature from 2000 onward mentioning video or multimedia learning. The additional criterion required was at least the use of one of the variables categorized under "usage" or "interaction," as described in the last section.

In Table 22.1, we introduce studies that have used such methods to explore learner use, engagement, and learning with videos. We have categorized the studies using the algorithm categories and application types noted on Figure 22.5.

In addition to the definition of each variable, the type of study or algorithm uses the same taxonomy presented earlier with studies that use the comparative approach or studies that apply data mining techniques classed using the schema in Figure 22.5. Notably, a "modelling" type, referring to work, has been added, which used the data and variables not to inform the learning and teaching per se, but to explain or describe the patterns of use and interaction with videos.

Moving from Usage to Engagement

As seen in the previous section, a number of studies investigate what learners do with videos. However, in order to characterize *engagement* in the context of learning and teaching, it is essential to consider what is meant by engagement with videos. For example, a learner could click on the play button of a video presented as part of a "flipped classroom" activity and then walk away to make coffee. The video would still be playing, and logging this activity as *usage*; however, the learner would be not be *engaged* with the activity. This poses a challenge when interpreting activity logs and makes the case for avoiding the "big wrench" approach mentioned earlier. The time spent on task is not simple to interpret in an ecologically valid setting; unlike in experimental conditions in which extraneous variables are controlled or monitored, real learning might occur in highly noisy conditions (for example, increasingly "on the go" from a mobile device on a busy commuter bus (Chen, Seilhamer, Bennett, & Bauer, 2015).

However, expedients made available via modern web technologies can partly circumvent this problem. For example, including in-video quizzes (IVQ) provides an opportunity to check whether learners have understood concepts in the video or how they perceive its effectiveness. These not only provide a "pulse" on engagement, but also a view on the effectiveness of the videos for learning. Most MOOC providers offer some form of IVQs that can be inserted at specific points.

Table 22.1. Summary of Related Work using Video Analytics Techniquesmat

Reference	Type of study or algorithm type	Usage			Interaction				Additional var.				
		Overall usage	Navigation	In-video behavior	Tagging/Pagging	Annotation	Social interaction	Learning design	Satisfaction	Perceived effect.	User information	Assessment	Content-related
Anusha & Shereen, 2014	Classification					v						v	
Avlonitis & Chorianopoulos, 2014	Correlation	v	v	v									v
Avlonitis, Karydis, & Sioutas, 2015	Correlation, Regression	v	v	v									v
Brooks, Epp, Logan, & Greer, 2011	Clustering	v	v	v						v			
Chen, Chen, Xu, March, & Benford, 2008	Comparison					v							v
Chorianopoulos, 2012	Comparison	v			o		v					v	v
Chorianopoulos, 2013	Comparison		v	v								v	v
Chorianopoulos, Giannakos, Chrisochoides, & Reed, 2014	Framework	v	v	v								v	v
Cobarzan & Schoeffmann, 2014	Comparison		v	v								v	v
Coleman, Seaton, & Chuang, 2015	Modelling, Classification	v										v	
Crockford & Agius, 2006	Comparison		v	v									v
de Konig, Tabbers, Rikers, & Paas, 2011	Comparison	v		v				v					v
Delen, Liew, & Willson, 2014	Comparison		v	v									v
Dufour, Toms, Lewis, & Baecker, 2005	Comparison	v	v	v					v				v
Gašević, Mirriahi, & Dawson, 2014	Comparison					v		v					
Giannakos, Chorianopoulos, & Chrisochoides 2014	Comparison		v	v						v		v	
Giannakos, Chorianopoulos, & Chrisochoides, 2015	Modelling, Comparison		v	v					v	v		v	v

Table 22.1 (cont.). Summary of Related Work using Video Analytics Techniquesmat

Reference	Type of study or algorithm type	Usage			Interaction				Additional var.				
		Overall usage	Navigation	In-video behavior	Tagging/Pagging	Annotation	Social interaction	Learning design	Satisfaction	Perceived effect.	User information	Assessment	Content-related
Giannakos, Jaccheri, & Krogstie, 2015	Correlation	v	v	v						v	v		
Gkonela & Chorianopoulus, 2012	Modelling	v	v	v			v				v		v
Grigoras, Charvillat & Douze, 2002	Modelling, Regression	v	v	v									v
Guo, Kim, & Rubin, 2014	Comparison	v						v				v	v
He, 2013	Correlation			v			v					v	
He, Grudin, & Gupta, 2000	Modelling, Clustering	v		v									
He, Sanocki, Gupta, & Grudin, 2000	Comparison	v						v				v	v
Ilioudi, Giannakos, & Chorianopoulos, 2013	Comparison						v		v			v	v
Kamahara, Nagamatsu, Fukuhara, Kaieda, & Ishii, 2009	Modelling		v	v									v
Kim, Guo, Cai, Li, Gajos, & Miller, 2014	Modelling, Comparison	v	v	v				v	v	v	v		v
Kim, Guo, Seaton, Mitros, Gajos, & Miller 2014	Modelling, Regression	v	v					v					v
Li, Gupta, Sanoki, He, & Rui, 2000	Modelling, Comparison	v	v	v					v	v			
Li, Kidzinski, Jermann, & Dillenbourg, 2015	Modelling, Regression	v	v	v						v			v
Li, Zhang, Hu, Zhu, Chen, Jiang, Deng, Guo, Faraco, Zhang, Han, Hua, & Liu, 2010	Modelling, Comparison	v									v		v
Lyons, Reysen, & Pierce, 2012	Comparison							v	v	v			v
Mirriahi & Dawson, 2013	Correlation	v				v						v	
Mirriahi, Liaqat, Dawson, & Gašević, 2016	Clustering	v	v	v		v							
Monserrat, Zhao, McGee, & Pendey, 2013	Comparison	v									v	v	v
Mu, 2010	Comparison	v				v		v			v		
Pardo, Mirriahi, Dawson, Zhao, Zhao, & Gašević, 2015	Correlation, Regression	v				v					v	v	
Phillips, Maor, Preston, & Cumming-Potvin, 2011	Comparison	v							v	v	v		
Risko, Foulsham, Dawson, & Kingstone, 2013	Modelling	v	v	v	o	v			v				
Ritzhaupt, Pastore & Davis, 2015	Correlation			v				v					
Samad & Hamid, 2015	Modelling, Comparison	v						v					
Schwan & Riempp, 2004	Comparison		v	v				v				v	
Shi, Fu, Chen, & Qu, 2014	Modelling, Comparison	v	v	v									
Sinha & Cassell, 2015	Modelling, Regression	v										v	
Song, Hong, Oakley, Cho, & Bianchi, 2015	Modelling	v		o							v		
Syeda-Mahmood & Poncelon, 2001	Clustering	v	v	v									v
Vondrick & Ramanan, 2011	Correlation	v						v					v
Weir, Kim, Gajos, & Miller, 2015	Comparison	v			v							v	v
Wen & Rose, 2014	Clustering	v											v
Wieling & Hofman, 2010	Correlation, Regression	v									v		v
Yu, Ma, Nahrstedt, & Zhang, 2003	Modelling, Clustering	v	v	v									v
Zahn, Barquero, & Schwan, 2004	Comparison	v	v	v							v	v	v
Zhang, Zhou, Briggs, & Nunamaker, 2006	Comparison	v						v			v	v	v
Zupancic & Horz, 2002	Comparison	v						v			v	v	v

NOTES: where 'o' is present, the attribution is subject to interpretation.
Overall usage refers to counts of activity; **Navigation** refers to browse, search and skip; **In-video behaviours** provides details of play, pause, back, forwards and speed change; **Tagging/Flagging** is the simple action of marking a time point which may contain a keyword (tagging) or visual indicator (flagging); **Annotation** refers to the ability to add text or other references to the videos; **Social Interaction** refers to the ability to share inputs or outputs with others; **Learning design** refers to specifically designed conditions, activities, or modes of presentation; **Satisfaction** and **Perceived effectiveness** are obtained via surveys or other in-line methods (i.e. quizzes); **User information** refers to the availability of user details (i.e. demographics); **Assessment** implies the presence of performance tests or grades; **Content-related** means that the input/outputs have a relevance for the content.

Giannakos et al. (2014) offered an interesting approach to studying in-video behaviours, testing the affordances of different types of implementation of videos and quiz combinations. The SocialSkip web application (http://www.socialskip.org) allows instructors to test different scenarios and see the results on students' navigation and performance. In this sense, Kozma's argument that media (and videos in this case) have defining characteristics interacting with the learner, the task characteristics are dependent on the instructional design that employs them and therefore shapes the type of engagement possible with the medium, which can be directly tested with analytics.

Gauging Learning from Usage and Engagement

There is evidence that some learners like the opportunities provided by video (Merkt, Weigand, Heier, &

Schwan, 2011) and, under certain conditions and with particular designs, videos lead to better learning (using achievement level and grades as proxies for learning) (Giannakos et al., 2014; Mirriahi & Dawson, 2013). Lab experiments have demonstrated that videos lead to better retention and recall (Mayer, Heiser, & Lonn, 2001), but the issue of transfer is a serious weakness in most studies. How do we go about demonstrating learning? If more active engagement with content facilitates deep learning, can videos provide this opportunity and, if so, under what conditions?

The discussion about whether it is possible to determine whether learning occurs based on the level of *engagement* with videos is a tricky one. Earlier we considered simple time on task as an ineffective way to measure engagement. In fact, engagement in learning and teaching can be characterized as having six dimensions (Figure 22.6): intellectual, emotional, behavioural, physical, social, and cultural. The dimensions most relevant when considering the use of videos are strictly intertwined with the nature of use and integration in the curriculum and, therefore, the type of data required to explore engagement for learning is dependent on the learning design and the technology available.

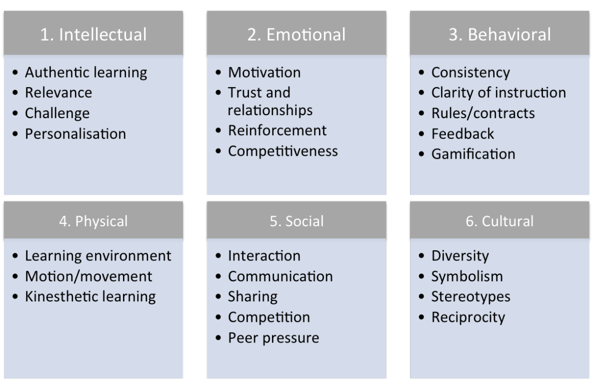


Figure 22.6. Elements of engagement in learning and teaching.

For example, a combination of analytics from learners' use of videos alongside surveys can help capture the metrics related to the intellectual, emotional, and cultural dimensions of learning such as whether they find the videos relevant or challenging and their motivation towards watching them. With the addition of IVQs, feedback and clarity of instruction (the behavioural dimension) can be explored. With the use and sharing of video tagging and annotation, the social dimension can be considered; if videos are used in the classroom, there is opportunity to understand the effects of the physical environment on learning. One of the fundamental problems is the inability to extricate the effect of videos on learning because the proxy of learning is often student performance or achievement demonstrated through assessments external to the video activity (with the exception of IVQs that could be used as summative quizzes assessing content directly related to a video). This poses a challenge for making appropriate judgements on the extent of learning achieved through engagement with videos. Yet, we can rely on the reported levels of satisfaction that learners provide as feedback with the usefulness or effectiveness of videos for their learning, providing a glimpse in their learning experience.

SUMMARY & FUTURE DIRECTIONS

The overview provided in this chapter is meant to introduce learning scientists, researchers, educators, and others interested in investigating the impact of videos on learning, use, and engagement to prior approaches that can be adapted to explore new questions and hypothesis.

The chapter has provided an overview of the types of potential videos used for educational purposes and the various ways they can be integrated into the curriculum (although by no means an exhaustive list). Data mining approaches, as one method of analyzing relevant data (alongside more traditional approaches) about learner use, engagement, and learning with videos is discussed with a short summary of approaches reported in recent studies as a starting point for interested readers to explore further as they wish.

This chapter has introduced video analytics and some applications showing how this approach can support the investigation and evaluation of learner engagement and learning with videos. Situating the concept of videos in the curriculum within multimedia learning provides a theoretical foundation for considering the ways in which multimedia (and videos) are included in learning and teaching and how they have historically been evaluated for their effectiveness. As we have seen, a single approach, or "big wrench," may not be as appropriate as a combination of methods and approaches.

Despite the considerable research accrued on the evaluation of the effectiveness of multimedia and videos for learning, many questions remain. Expanding on the studies and strategies to date and leveraging the growing body of data being captured by video technologies provides an opportunity to investigate a milieu of questions not limited to the following:

1. How do learners use, engage with, and learn from different types of videos (e.g., reflection vs. lecture)? A mixed methods approach consisting of video analytics, learner feedback, and instructor reflections and in variety of curriculum or learning

design contexts would be useful here.

2. What types of interventions can be applied to either a) inform instructors of changes required to video content or how it is integrated in the curriculum or b) inform learners of learning strategies to better engage with the video content or associated activities? Once such interventions are identified, their effectiveness would need to be explored.
3. Rather than relying on proxies of learning (e.g., assessment scores or final marks), how can learning with or from videos be more accurately nuanced?
4. How can data related to video content be better mined in order to explore how it affects learner use and engagement patterns?
5. What are the most effective algorithms for this domain? Is it possible to use some of the models generated to inform instructional design and provide further opportunities to improve learning and the student experience?

REFERENCES

Antani, S., Kasturi, R., & Jain, R. (2002). A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video. *Pattern Recognition*, *35*(4), 945–965. http://doi.org/10.1016/S0031-3203(01)00086-3

Anusha, V., & Shereen, J. (2014). Multiple lecture video annotation and conducting quiz using random tree classification. *International Journal of Engineering Trends and Technology*, 8(10), 522–525.

Atkins, M. J. (1993). Theories of learning and multimedia applications: An overview. *Research Papers in Education*, 8(2), 251–271. http://doi.org/10.1080/0267152930080207

Avlonitis, M., & Chorianopoulos, K. (2014). Video pulses: User-based modeling of interesting video segments. *Advances in Multimedia*, *2014*. http://doi.org/10.1155/2014/712589

Avlonitis, M., Karydis, I., & Sioutas, S. (2015). Early prediction in collective intelligence on video users' activity. *Information Sciences*, 298, 315–329. http://doi.org/10.1016/j.ins.2014.11.039

Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne social desirability scale. *Educational and Psychological Measurement*, 62(4), 570–589. http://doi.org/10.1177/0013164402062004003

Bloom, B. S. (1968). Learning for Mastery: Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation Comment*, *1*(2), n2.

Brooks, C., Epp, C. D., Logan, G., & Greer, J. (2011). The who, what, when, and why of lecture capture. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 86–92). New York: ACM.

Brunelli, R., Mich, O., & Modena, C. M. (1999). A survey on the automatic indexing of video data. *Journal of Visual Communication and Image Representation*, 10(2), 78–112. http://doi.org/10.1006/jvci.1997.0404

Chen, B., Seilhamer, R., Bennett, L., & Bauer, S. (2015, June 22). Students' mobile learning practices in higher education: A multi-year study. *Educause Review*. http://er.educause.edu/articles/2015/6/students-mobile-learning-practices-in-higher-education-a-multiyear-study

Chen, L., Chen, G.-C., Xu, C.-Z., March, J., & Benford, S. (2008). EmoPlayer: A media player for video clips with affective annotations. *Interacting with Computers*, 20(1), 17–28. http://doi.org/10.1016/j.intcom.2007.06.003

Chorianopoulos, K. (2012). Crowdsourcing user interactions with the video player. *Proceedings of the 18th Brazilian Symposium on Multimedia and the Web* (WebMedia '12), 15–18 October 2012, São Paulo, Brazil (pp. 13–16). New York: ACM. http://doi.org/10.1145/2382636.2382642

Chorianopoulos, K. (2013). Collective intelligence within web video. *Human-Centric Computing and Information Sciences*, 3(1), 1–16. http://doi.org/10.1186/2192-1962-3-10

Chorianopoulos, K., Giannakos, M. N., Chrisochoides, N., & Reed, S. (2014). Open service for video learning analytics. *Proceedings of the 14th IEEE International Conference on Advanced Learning Technologies* (ICALT

2014), 7–10 July 2014, Athens, Greece (pp. 28–30). http://doi.org/10.1109/ICALT.2014.19

Clark, R. E. (1983). Reconsidering research on learning from media. *Review of Educational Research*, 53(4), 445–459. http://doi.org/10.3102/00346543053004445

Clark, R. E. (1994). Media and method. *Educational Technology Research and Development*, 42(3), 7–10. http://doi.org/10.1007/BF02298090

Cobârzan, C., & Schoeffmann, K. (2014). How do users search with basic HTML5 video players? In C. Gurrin, F. Hopfgartner, W. Hurst, H. Johansen, H. Lee, & N. O'Connor (Eds.), *MultiMedia Modeling* (pp. 109–120). Springer. http://link.springer.com.wwwproxy0.library.unsw.edu.au/chapter/10.1007/978-3-319-04114-8_10

Colasante, M. (2010). Future-focused learning via online anchored discussion, connecting learners with digital artefacts, other learners, and teachers. *Proceedings of the 27th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Curriculum, Technology & Transformation for an Unknown Future* (ASCILITE 2010), 5–8 December 2010, Sydney, Australia (pp. 211–221). ASCILITE.

Colasante, M. (2011). Using video annotation to reflect on and evaluate physical education pre-service teaching practice. *Australasian Journal of Educational Technology*, 27(1), 66–88.

Coleman, C. A., Seaton, D. T., & Chuang, I. (2015). Probabilistic use cases: Discovering behavioral patterns for predicting certification. *Proceedings of the 2nd ACM conference on Learning@Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 141–148). New York: ACM. http://doi.org/10.1145/2724660.2724662

Conole, G. (2013). MOOCs as disruptive technologies: Strategies for enhancing the learner experience and quality of MOOCs. *Revista de Educación a Distancia (RED)*, 39. http://www.um.es/ead/red/39/conole.pdf

Crockford, C., & Agius, H. (2006). An empirical investigation into user navigation of digital video using the VCR-like control set. *International Journal of Human–Computer Studies*, 64(4), 340–355. http://doi.org/10.1016/j.ijhcs.2005.08.012

Daniel, R. (2001). Self-assessment in performance. *British Journal of Music Education*, 18(3). http://doi.org/10.1017/S0265051701000316

Dawson, S., Macfadyen, L., Evan, F. R., Foulsham, T., & Kingstone, A. (2012). Using technology to encourage self-directed learning: The Collaborative Lecture Annotation System (CLAS). *Proceedings of the 29th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education* (ASCILITE 2012), 25–28 October, Wellington, New Zealand (pp. XXX–XXX). ASCILITE. http://www.ascilite.org/conferences/Wellington12/2012/images/custom/dawson,_shane_-_using_technology.pdf

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior*, 27(1), 41–45. http://doi.org/10.1016/j.chb.2010.05.010

Delen, E., Liew, J., & Willson, V. (2014). Effects of interactivity and instructional scaffolding on learning: Self-regulation in online video-based environments. *Computers & Education*, 78, 312–320. http://doi.org/10.1016/j.compedu.2014.06.018

Diwanji, P., Simon, B. P., Marki, M., Korkut, S., & Dornberger, R. (2014). Success factors of online learning videos. *Proceedings of the International Conference on Interactive Mobile Communication Technologies and Learning* (IMCL 2014), 13–14 November 2014, Thessaloniki, Greece (pp. 125–132). IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7011119

Dufour, C., Toms, E. G., Lewis, J., & Baecker, R. (2005). User strategies for handling information tasks in webcasts. *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '05), 2–7 April 2005, Portland, OR, USA (pp. 1343–1346). New York: ACM. http://doi.org/10.1145/1056808.1056912

El Samad, A., & Hamid, O. H. (2015). The role of socio-economic disparities in varying the viewing behavior of e-learners. *Proceedings of the 5th International Conference on Digital Information and Communication Technology and its Applications* (DICTAP) (pp. 74–79). https://doi.org/10.1109/DICTAP.2015.7113174

Fegade, M. A., & Dalal, V. (2014). A survey on content based video retrieval. *International Journal of Engineering and Computer Science*, 3(7), 7271–7279.

Gagne, R. M. (1965). *The conditions of learning*. Holt, Rinehart & Winston.

Gašević, D., Mirriahi, N., & Dawson, S. (2014). Analytics of the effects of video use and instruction to support reflective learning. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 123–132). New York: ACM. http://doi.org/10.1145/2567574.2567590

Giannakos, M. N. (2013). Exploring the video-based learning research: A review of the literature: Colloquium. *British Journal of Educational Technology*, 44(6), E191–E195. http://doi.org/10.1111/bjet.12070

Giannakos, M. N., Chorianopoulos, K., & Chrisochoides, N. (2014). Collecting and making sense of video learning analytics. *Proceedings of the 2014 IEEE Frontiers in Education Conference* (FIE 2014), 22–25 October 2014, Madrid, Spain. IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7044485

Giannakos, M. N., Chorianopoulos, K., & Chrisochoides, N. (2015). Making sense of video analytics: Lessons learned from clickstream interactions, attitudes, and learning outcome in a video-assisted course. *The International Review of Research in Open and Distributed Learning*, 16(1). http://www.irrodl.org/index.php/irrodl/article/view/1976

Giannakos, M. N., Jaccheri, L., & Krogstie, J. (2015). Exploring the relationship between video lecture usage patterns and students' attitudes: Usage patterns on video lectures. *British Journal of Educational Technology*, 47(6), 1259–1275. http://doi.org/10.1111/bjet.12313

Gkonela, C., & Chorianopoulos, K. (2012). VideoSkip: Event detection in social web videos with an implicit user heuristic. *Multimedia Tools and Applications*, 69(2), 383–396. http://doi.org/10.1007/s11042-012-1016-1

Gonyea, R. M. (2005). Self-reported data in institutional research: Review and recommendations. *New Directions for Institutional Research*, 2005(127), 73–89.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3), 42–57.

Grigoras, R., Charvillat, V., & Douze, M. (2002). Optimizing hypervideo navigation using a Markov decision process approach. *Proceedings of the 10th ACM International Conference on Multimedia* (MULTIMEDIA '02), 1–6 December 2002, Juan-les-Pins, France (pp. 39–48). New York: ACM. http://doi.org/10.1145/641007.641014

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the 1st ACM conference on Learning@Scale* (L@S 2014), 4–5 March 2014, Atlanta, Georgia, USA (pp. 41–50). New York: ACM. http://doi.org/10.1145/2556325.2566239

Guskey, T. R., & Good, T. L. (2009). Mastery learning. In T. L. Good (Ed.), *21st Century Education: A Reference Handbook*, vol. 1 (pp. 194–202). Thousand Oaks, CA: Sage.

He, W. (2013). Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, 29(1), 90–102. http://doi.org/10.1016/j.chb.2012.07.020

He, L., Grudin, J., & Gupta, A. (2000). Designing presentations for on-demand viewing. *Proceedings of the 2000 Conference on Computer Supported Cooperative Work* (CSCW '00), 2–6 December 2000, Philadelphia, PA, USA (pp. 127–134). New York: ACM. http://doi.org/10.1145/358916.358983

He, L., Sanocki, E., Gupta, A., & Grudin, J. (2000). Comparing presentation summaries: Slides vs. reading vs. listening. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI 2000), 1–6 April 2000, The Hague, Netherlands (pp. 177–184). New York: ACM. http://doi.org/10.1145/332040.332427

Hulsman, R. L., Harmsen, A. B., & Fabriek, M. (2009). Reflective teaching of medical communication skills with DiViDU: Assessing the level of student reflection on recorded consultations with simulated patients. *Patient Education and Counseling*, 74(2), 142–149. http://doi.org/10.1016/j.pec.2008.10.009

Ilioudi, C., Giannakos, M. N., & Chorianopoulos, K. (2013). Investigating differences among the commonly used video lecture styles. *Proceedings of the Workshop on Analytics on Video-Based Learning* (WAVe 2013), 8 April 2013, Leuven, Belgium (pp. 21–26). http://ceur-ws.org/Vol-983/WAVe2013-Proceedings.pdf

Johnson, T. (2011, July 22). A few notes from usability testing: Video tutorials get watched, text gets skipped. I'd Rather Be Writing [Blog]. http://idratherbewriting.com/2011/07/22/a-few-notes-from-usability-testing-video-tutorials-get-watched-text-gets-skipped/

Joy, E. H., & Garcia, F. E. (2000). Measuring learning effectiveness: A new look at no-significant-difference findings. *Journal of Asynchronous Learning Networks*, 4(1), 33–39.

Juhlin, O., Zoric, G., Engström, A., & Reponen, E. (2014). Video interaction: A research agenda. *Personal and Ubiquitous Computing*, 18(3), 685–692. http://doi.org/10.1007/s00779-013-0705-8

Kamahara, J., Nagamatsu, T., Fukuhara, Y., Kaieda, Y., & Ishii, Y. (2009). Method for identifying task hardships by analyzing operational logs of instruction videos. In T.-S. Chua, Y. Kompatsiaris, B. Mérialdo, W. Haas, G. Thallinger, & W. Bailer (Eds.), *Semantic Multimedia* (pp. 161–164). Springer. http://link.springer.com.www-proxy0.library.unsw.edu.au/chapter/10.1007/978-3-642-10543-2_16

Keller, F. S. (1967). Engineering personalized instruction in the classroom. *Revista Interamericana de Psicologia*, *1*(3), 144–156.

Kim, J., Guo, P. J., Cai, C. J., Li, S.-W. (Daniel), Gajos, K. Z., & Miller, R. C. (2014). Data-driven Interaction Techniques for Improving Navigation of Educational Videos. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (pp. 563–572). New York, NY, USA: ACM. https://doi.org/10.1145/2642918.2647389

Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller, R. C. (2014). Understanding in-video dropouts and interaction peaks in online lecture videos. *Proceedings of the 1st ACM Conference on Learning @ Scale* (L@S 2014), 4–5 March 2014, Atlanta, GA, USA (pp. 31–40). New York: ACM. http://doi.org/10.1145/2556325.2566239

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, *41*(2), 75–86. http://doi.org/10.1207/s15326985ep4102_1

Kolowich, S. (2013, March 18). The professors who make the MOOCs. *The Chronicle of Higher Education*, 25. http://www.chronicle.com/article/The-Professors-Behind-the-MOOC/137905/

Kozma, R. B. (1991). Learning with media. *Review of Educational Research*, 61(2), 179–211. http://doi.org/10.3102/00346543061002179

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42(2), 7–19. http://doi.org/10.1007/BF02299087

Kulik, C.-L. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research*, 60(2), 265–299. http://doi.org/10.3102/00346543060002265

Lee, H. S., & Anderson, J. R. (2013). Student learning: What has instruction got to do with it? *Annual Review of Psychology*, 64(1), 445–469. http://doi.org/10.1146/annurev-psych-113011-143833

Li, F. C., Gupta, A., Sanocki, E., He, L., & Rui, Y. (2000). Browsing digital video. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI 2000), 1–6 April 2000, The Hague, Netherlands (pp. 169–176). New York: ACM. http://doi.org/10.1145/332040.332425

Li, N., Kidzinski, L., Jermann, P., & Dillenbourg, P. (2015). How do in-video interactions reflect perceived video difficulty? *Proceedings of the 3rd European MOOCs Stakeholder Summit*, 18–20 May 2015, Mons, Belgium (pp. 112–121). PAU Education. http://infoscience.epfl.ch/record/207968

Li, K., T. Zhang, X. Hu, D. Zhu, H. Chen, X. Jiang, F. Deng, J. Lv, C. C. Faraco, and D. Zhang. 2010. “Human-Friendly Attention Models for Video Summarization.” In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction* (pp. 27:1–27:8). New York: ACM. http://doi.org/10.1145/1891903.1891938

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior*, 28(1), 181–186. http://doi.org/10.1016/j.chb.2011.08.025

Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of Massive Open Online Courses (MOOCs). *Computers & Education*, 80, 77–83. http://doi.org/10.1016/j.compedu.2014.08.005

Mayer, R. E. (2009). *Multimedia learning*. Cambridge, UK: Cambridge University Press.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93(1), 187–198. http://doi.org/10.1037/0022-0663.93.1.187

McKeachie, W. J. (1974). Instructional psychology. *Annual Review of Psychology*, *25*(1), 161–193. http://doi.org/10.1146/annurev.ps.25.020174.001113

Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, *21*(6), 687–704. http://doi.org/10.1016/j.learninstruc.2011.03.004

Mirriahi, N., & Dawson, S. (2013). The pairing of lecture recording data with assessment scores: A method of discovering pedagogical impact. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 180–184). New York: ACM. http://dl.acm.org/citation.cfm?id=2460331

Mirriahi, N., Liaqat, D., Dawson, S., & Gašević, D. (2016). Uncovering student learning profiles with a video annotation tool: Reflective learning with and without instructional norms. *Educational Technology Research and Development*, *64*(6), 1083–1106. http://doi.org/10.1007/s11423-016-9449-2

Monserrat, T.-J. K. P., Zhao, S., McGee, K., & Pandey, A. V. (2013). NoteVideo: Facilitating navigation of Blackboard-style lecture videos. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '13), 27 April–2 May 2013, Paris, France (pp. 1139–1148). New York: ACM. http://doi.org/10.1145/2470654.2466147

Mu, X. (2010). Towards effective video annotation: An approach to automatically link notes with video content. *Computers & Education*, *55*(4), 1752–1763. http://doi.org/10.1016/j.compedu.2010.07.021

Oblinger, D. G., & Hawkins, B. L. (2006). The myth about no significant difference. *EDUCAUSE Review*, *41*(6), 14–15.

Owston, R., Lupshenyuk, D., & Wideman, H. (2011). Lecture capture in large undergraduate classes: Student perceptions and academic performance. *The Internet and Higher Education*, 14(4), 262–268. http://doi.org/10.1016/j.iheduc.2011.05.006

Palincsar, A. S. (1998). Social constructivist perspectives on teaching and learning. *Annual Review of Psychology*, 49(1), 345–375. http://doi.org/10.1146/annurev.psych.49.1.345

Pardo, A., Mirriahi, N., Dawson, S., Zhao, Y., Zhao, A., & Gašević, D. (2015). Identifying learning strategies associated with active use of video annotation software. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 255–259). New York: ACM Press. http://doi.org/10.1145/2723576.2723611

Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, *41*(4, Part 1), 1432–1462. http://doi.org/10.1016/j.eswa.2013.08.042

Phillips, R., Maor, D., Cumming-Potvin, W., Roberts, P., Herrington, J., Preston, G., … Perry, L. (2011). Learning analytics and study behaviour: A pilot study. In G. Williams, P. Statham, N. Brown, & B. Cleland (Eds.), *Proceedings of the 28th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education: Changing Demands, Changing Directions* (ASCILITE 2011), 4–7 December 2011, Hobart, Tasmania, Australia (pp. 997–1007). ASCILITE. http://researchrepository.murdoch.edu.au/6751/

Reeves, T. C. (1986). Research and evaluation models for the study of interactive video. *Journal of Computer-Based Instruction*, 13(4), 102–106.

Reeves, T. C. (1991). Ten commandments for the evaluation of interactive multimedia in higher education. *Journal of Computing in Higher Education*, 2(2), 84–113. http://doi.org/10.1007/BF02941590

Risko, E. F., Foulsham, T., Dawson, S., & Kingstone, A. (2013). The collaborative lecture annotation system (CLAS): A new TOOL for distributed learning. *IEEE Transactions on Learning Technologies*, 6(1), 4–13. http://doi.org/10.1109/TLT.2012.15

Ritzhaupt, A. D., Pastore, R., & Davis, R. (2015). Effects of captions and time-compressed video on learner performance and satisfaction. *Computers in Human Behavior*, 45, 222–227. http://doi.org/10.1016/j.chb.2014.12.020

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146.

Romero, C., & Ventura, S. (2010). Educational data mining: A review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 40(6), 601–618. https://doi.org/10.1109/TSMCC.2010.2053532

Russell, T. L. (1999). The no significant difference phenomenon: A comparative research annotated bibliography on technology for distance education: As reported in 355 research reports, summaries and papers. North Carolina State University.

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction*, 14(3), 293–305. http://doi.org/10.1016/j.learninstruc.2004.06.005

Shi, C., Fu, S., Chen, Q., & Qu, H. (2014). VisMOOC: Visualizing video clickstream data from massive open online courses. *Proceedings of the 2014 IEEE Conference on Visual Analytics Science and Technology* (VAST 2014), 9–14 November 2014, Paris, France (pp. 277–278). IEEE. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7042528

Sinha, T., & Cassell, J. (2015). Connecting the dots: Predicting student grade sequences from Bursty MOOC interactions over time. *Proceedings of the 2nd ACM Conference on Learning@Scale* (L@S 2015), 14–18 March 2015, Vancouver, BC, Canada (pp. 249–252). New York: ACM. http://doi.org/10.1145/2724660.2728669

Skinner, F. B. (1950). Are theories of learning necessary? *Psychological Review*, 57(4), 193–216. http://doi.org/10.1037/h0054367

Song, S., Hong, J., Oakley, I., Cho, J. D., & Bianchi, A. (2015). Automatically adjusting the speed of e-learning videos. *CHI 33rd Conference on Human Factors in Computing Systems: Extended Abstracts* (CHI EA '15), 18–23 April 2015, Seoul, Republic of Korea (pp. 1451–1456). New York: ACM. http://doi.org/10.1145/2702613.2732711

Syeda-Mahmood, T., & Ponceleon, D. (2001). Learning video browsing behavior and its application in the generation of video previews. *Proceedings of the 9th ACM International Conference on Multimedia* (MULTIMEDIA '01), 30 September–5 October 2001, Ottawa, ON, Canada (pp. 119–128). New York: ACM. http://doi.org/10.1145/500141.500161

Tennyson, R. D. (1994). The big wrench vs. integrated approaches: The great media debate. *Educational Technology Research and Development*, 42(3), 15–28. http://doi.org/10.1007/BF02298092

Vondrick, C., & Ramanan, D. (2011). Video annotation and tracking with active learning. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, K. Q. Weinberger (Eds.), *Advances in Neural Information Processing Systems* 24 (NIPS 2011), 12–17 December 2011, Granada, Spain (pp. 28–36). http://papers.nips.cc/paper/4233-video-annotation-and-tracking-with-active-learning

Weir, S., Kim, J., Gajos, K. Z., & Miller, R. C. (2015). Learnersourcing subgoal labels for how-to videos. *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (CSCW '15), 14–18 March 2015, Vancouver, BC, Canada (pp. 405–416). New York: ACM. http://doi.org/10.1145/2675133.2675219

Wen, M., & Rosé, C. P. (2014). Identifying latent study habits by mining learner behavior patterns in massive open online courses. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management* (CIKM '14), 3–7 November 2014, Shanghai, China (pp. 1983–1986). New York: ACM. http://doi.org/10.1145/2661829.2662033

Wieling, M. B., & Hofman, W. H. A. (2010). The impact of online video lecture recordings and automated feedback on student performance. *Computers & Education*, 54(4), 992–998. http://doi.org/10.1016/j.compedu.2009.10.002

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*, 27(4), 551–572.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Burlington, MA: Morgan Kaufmann Publishers.

Yousef, A. M. F., Chatti, M. A., & Schroeder, U. (2014). Video-based learning: A critical analysis of the research published in 2003–2013 and future visions. *Proceedings of the 6th International Conference on Mobile, Hybrid, and On-line Learning* (ThinkMind/eLmL 2014), 23–27 March 2014, Barcelona, Spain (pp. 112–119). http://www.thinkmind.org/index.php?view=article&articleid=elml_2014_5_30_50050

Yu, B., Ma, W.-Y., Nahrstedt, K., & Zhang, H.-J. (2003). Video summarization based on user log enhanced link analysis. *Proceedings of the 11th ACM International Conference on Multimedia* (MULTIMEDIA '03), 2–8 November 2003, Berkeley, CA, USA (pp. 382–391). New York: ACM. http://doi.org/10.1145/957013.957095

Zahn, C., Barquero, B., & Schwan, S. (2004). Learning with hyperlinked videos: Design criteria and efficient strategies for using audiovisual hypermedia. *Learning and Instruction*, 14(3), 275–291. http://doi.org/10.1016/j.learninstruc.2004.06.004

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker Jr., J. F. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management*, 43(1), 15–27. http://doi.org/10.1016/j.im.2005.01.004

Zupancic, B., & Horz, H. (2002). Lecture recording and its use in a traditional university course. In *Proceedings of the 7th Annual Conference on Innovation and Technology in Computer Science Education* (ITiCSE '02), 24–28 June 2002, Aarhus, Denmark (pp. 24–28). New York: ACM. http://doi.org/10.1145/544414.544424

Chapter 23: Learning and Work: Professional Learning Analytics

Allison Littlejohn

Institute of Educational Technology, Open University, United Kingdom
DOI: 10.18608/hla17.023

ABSTRACT

Learning for work takes various forms, from formal training to informal learning through work activities. In many work settings, professionals collaborate via networked environments leaving various forms of digital traces and "clickstream" data. These data can be exploited through learning analytics (LA) to make both formal and informal learning processes traceable and visible to support professionals with their learning. This chapter examines the state-of-the-art in professional learning analytics (PLA) by considering how professionals learn, putting forward a vision for PLA, and analyzing examples of analytics in action in professional settings. LA can address affective and motivational learning issues as well as technical and practical expertise; it can intelligently align individual learning activities with organizational learning goals. PLA is set to form a foundation for future learning and work.

Keywords: Professional learning, formal training, informal learning

Professional learning is a critical component of ongoing improvement, innovation, and adoption of new practices for work (Boud & Garrick, 1999; Fuller et al., 2003; Engeström, 2008). In an uncertain business environment, organizations must be able to learn continuously in order to deal with continual change (IBM, 2008). Learning for work takes different forms, ranging from formal training to discussions with colleagues to informal learning through work activities (Eraut, 2004; Fuller et al., 2003). These actions can be conceived of as different learning contexts producing a variety of data that can be used to improve professional learning and development (Billett, 2004; Littlejohn, Milligan, & Margaryan, 2012). In contemporary workplaces, professionals tend to collaborate via networked environments, using digital resources, leaving various forms of digital traces and "clickstream" data. Analysis of these different types of data potentially provides a powerful means of improving operational effectiveness by enhancing and supporting the various ways professionals learn and adapt.

For some years now, employers have been aware of the potential of learning analytics to support and enhance professional learning (Buckingham Shum & Ferguson, 2012). Learning analytics (LA) is an emerging methodological and multidisciplinary research area aimed at "the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens & Long, 2011, p. 34).

The vision for LA in education was of multifaceted systems that could leverage data and adaptive models of pedagogy to support learning (Baker, 2016; Berendt, Vuorikari, Littlejohn, & Margaryan, 2014). These systems would mine the massive amounts of data generated as a by-product of digital learning activity to support learners in achieving their goals (Ferguson, 2012). However, the systems developed for use in university education have been much simpler, focusing on economic concerns associated with higher education cost and impact in terms of learner outcomes (Nistor, Derntl, & Klamma, 2015; HEC Report, 2016). Many LA systems are based on predictive models that analyze individual learner profiles to forecast whether a learner is "at risk of dropping out" (Siemens & Long, 2011, p. 36; Wolff, Zdrahal, Nikolov, & Pantucek, 2013; Berendt et al. 2014; Nistor et al., 2015). These data are then presented to learners or teachers using a variety of dashboards. Current research is focusing on the actions taken to follow up on this feedback (Rienties et al., 2016).

In educational settings, learning is focused on course objectives. However, in organizational settings learning processes have to be aligned with organizational and project goals (Kimmerle, Cress, & Held, 2010). Learning tends to be planned around annual performance review processes, usually overseen by a Human Resource department. This type of system works well in organizations where large groups have standard work tasks and plan similar development activities.

In many organizations, however, job roles are becoming specialized, requiring unique and personalized development planning. In these circumstances, the "top down" planning models, where goals and priorities are planned and sequenced from the outset, may not be effective. Some organizations are shifting from top-down and individualized development planning using enterprise systems to adaptive and collaborative activity planning based around grassroots use of technologies, shifting towards "smart" or "agile" development planning where project teams have discretion to change the direction of the project over time (Clow, 2013). This means that development goals cannot be planned at the beginning of each development cycle; new and emerging priorities arise as each project unfolds. Agile planning systems require adaptive, just-in-time learning where people acquire the knowledge they need for new work tasks as the tasks emerge. However, this means that professionals have to be able to plan and self-regulate their own learning and development, changing their learning priorities as their work tasks evolve (Littlejohn et al., 2012).

HOW PROFESSIONALS LEARN

Learning in educational settings tends to focus around individual learner outcomes and explicit pedagogical models. Professional learning, however, is driven by the demands of work tasks and is interwoven with work processes (Eraut, 2000). By "professional learning," I mean the activities professionals engage in to stimulate their thinking and professional knowledge, to improve work performance and to ensure that practice is informed and up-to-date (Littlejohn & Margaryan, 2013, p. 2).

Professionals themselves tend to think of learning in terms of training or formal learning (Eraut, 2000). Yet, there is a growing body of evidence that professional learning is more effective when integrated with work tasks (see, for example, Collin, 2008; Tynjälä, 2008; Fuller & Unwin, 2004; Eraut, 2004). This type of learning is difficult to distinguish from everyday work tasks, so professionals may not recognize instances of learning (Argyris & Schön, 1974; Engeström, 1999).

Eraut's (2004) work in particular foregrounds the importance of on-the-job learning, broadly describing professional learning as "intentional" and planned or "unintentional" and opportune. According to Tynjälä (2008) intentional learning may be pre-planned and structured as formal learning, for example degree programmes, classroom training, practical workshops, coaching or mentoring; other forms are less easy to recognize, for example asking a colleague for help or watching an expert perform a task. Learning can result as an "unintended" consequence of work activity (Eraut, 2000). A manager in finance organization might improve his inter-cultural competencies over time as new colleagues from branches around the world join his team (Littlejohn & Hood, 2016). Professionals may be unaware of this sort of experiential learning until they reflected on how their practice has evolved over time. These different forms of professional learning are illustrated in the typology in Figure 23.1.

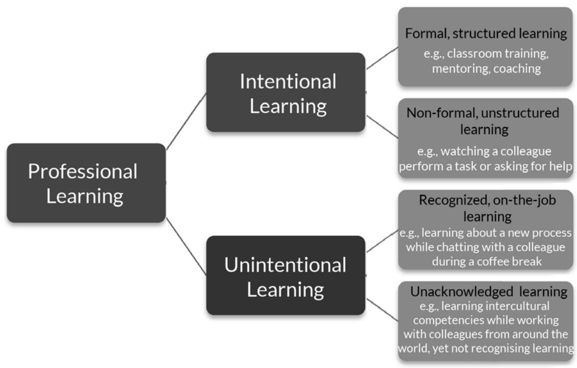


Figure 23.1. Typology of professional learning, informed by Eraut (2000, 2004).

The different approaches to learning illustrated in Figure 23.1 facilitate development of different types of knowledge (Tynjälä & Gijbels, 2012; Littlejohn & Hood, 2016). Education and training tend to focus on learning theoretical and practical knowledge, while coaching and mentoring allow opportunities to learn sociocultural and self-regulative knowledge, for example. All these knowledge types are critical for the adoption of new practices for work. Change in practice requires the construction of conceptual and practical knowledge as well as the development of sociocultural and self-regulative knowledge (Eraut, 2007). Construction of multiple types of knowledge is most readily achieved through a combination of formal (structured, pre-planned) learning activities with informal (unstructured, on-the-job) learning (Harteis & Billett, 2008). As such, workplace learning operates as a reciprocal process (Billett, 2004) shaped by the affordances of a specific workplace, together with an individual's ability and motivation to engage with what is afforded (Billett, 2004; Fuller & Unwin, 2004).

Learning processes for work are more dynamic than in educational settings; Informal learning activities are spontaneous and mostly invisible to others. This presents challenges and opportunities for the field of LA.

A VISION FOR PROFESSIONAL LEARNING ANALYTICS

An underlying vision for LA in professional contexts is to make both formal and informal learning processes traceable and more explicit in order to connect each professional with the knowledge they need (Littlejohn et al., 2012; de Laat & Schreurs, 2013). This vision is based on a system of mutual support through which each professional connects with and contributes to the collective knowledge by *connecting* with people and networks to find relevant knowledge and experiences; *consuming* or using this knowledge and, in the process, *creating* new knowledge that is contributed back to the collective (Milligan, Littlejohn, & Margaryan, 2014). These actions create a common capital through re-usable knowledge via the selective accumulation of shared by-products of individual activities motivated, initially, by personal utility (Convertino, Grasso, DiMicco, De Michelis, & Chi, 2010, p. 15). These actions would be supported by a set of algorithms, data mining mechanisms, and analytics that create a "common capital through re-usable knowledge via the selective accumulation of shared by-products of individual activities motivated, initially, by personal utility" (Convertino et al., 2010, p. 15).

Professional learning is influenced by the learner's internal motivation and personal agency in connecting to and interacting with the collective knowledge and their environment (Littlejohn & Hood, 2016), therefore there are two critical components to this vision. First, to ensure personal agency it is critical that professionals have the ability to self-regulate their learning. Second, to trigger motivation, learning (and learning systems) should be integrated with, rather than separate from, work practices. In moving towards this vision, a range of approaches to Professional LA have been developed over the past few years.

Analytics in Action

LA is a multidisciplinary area using ideas from learning science, computer science, information science, educational data mining, knowledge management highlighting, and, more recently, artificial intelligence (Gillani & Enyon, 2014). Emerging areas of analytics make use of complex datasets containing multiple data types such as discourse data, learner disposition data, and biometrics (Siemens & Long, 2011). Techniques used in LA include discourse analysis, where learners discussions and actions provide opportunity for helpful interventions (Gillani & Enyon, 2014); semantic analysis, tracing the relationship between learners and learning (Wen, Yang, & Rosé, 2014), learner disposition analytics, identifying affective characteristics associated with learning (Buckingham Shum & Deakin Crick, 2012) and content analytics, including recommender systems that filter and deliver content based on tags and ratings supplied by learners. These techniques are useful for encapsulating the complex factors that influence how professionals learn.

Diverse approaches to LA have been field-tested in various professional settings. Some methods capitalize on the data generated as a by-product of learning in digital systems. Others use new approaches, for example social learning analytics (SLA) that examine how individuals and groups learn and develop new knowledge. These methods and systems capitalize on new forms of organization, different feedback formats, and the numerous ways people and the resources they require for their learning and work can be brought together. Many are in the early stages of development and this section examines different approaches and their impact on learning and performance.

Accelerating Just-in-Time Learning

Some approaches to PLA are aimed towards embedding agile approaches to professional learning in work settings. Many organizations recognize that training is not effective if professionals learn a new process then do not use their new knowledge and embed it within their practice. Recognizing the importance of enabling people to learn new expertise at the point of need, organizations have been seeking ways to capture and disseminate expertise.

Wearable Experience for Knowledge Intensive Training (WEKIT)[1] is exploring if and how data generated through smart Wearable Technology can capture expertise and disseminate the know-how to inexperienced professionals at the point of need. WEKIT is based on a three-stage process: mapping skill development pathways, capturing and codifying expertise, and making the expertise available to novices at the point of need. In the first stage, a community of professionals and stakeholders[2] map out a recognized skill development pathways for industry. In the second stage, a group of software developers use the pathway templates to develop technology tools to support novices in learning new procedural knowledge – for example how to turn on (or off) a specialist valve. Finally, the expertise is transmitted to the novice via an augmented visual interface. Tools such as head-mounted digital displays allow the novice to see the valve overlaid with instructions on how to switch it on safely. Through

[1] https://wekit-community.org/
[2] the WEKIT.club

wearable and visual devices, the system directs each professional's attention to where it is most needed, based on an analysis of user needs. The system aims to make informal learning processes traceable and recognizable so that novices can rapidly develop expertise. In this way, learning can be agile, as the need arises.

The WEKIT project commenced in December 2015 and evaluations of the effectiveness of the approach have yet to be published. However, the three key steps in the transfer of expertise in the WEKIT methodology all have risks associated with them. First, expertise development pathways are difficult to model. Experts are involved in building the pathways and algorithms to support expertise development in an attempt to capture and codify the expertise accurately. However, it is difficult for an expert to understand the optimal learning pathway that will enable each novice's expertise development, since this depends on the novice's prior experience. Second, not all expertise can be codified. Augmented visual interfaces and collaborative digital interfaces can help with expertise development, but tacit expertise, such as the "gut feeling" that a piece of equipment is operating optimally, takes time to be developed. Thirdly, the novice has to be actively involved in learning new expertise, rather than simply following instructions. This is particularly relevant in work settings where task outcomes are difficult to predict, such as knowledge work. Learning in these situations is most effective when integrated with work tasks. Therefore, an emerging trend is to embed PLA within work-integrated systems: platforms that support experts and novices in co-working, smart systems or augmented reality environments, such as those described earlier.

Exploiting Organizational Networks

By tapping into informal professional networks, people can achieve the kind of agile learning described by Clow (2013). This type of self-governing, bottom-up approach to professional development requires a deep understanding of how and where professionals interact and exchange ideas about their work. Making informal learning practices and networks visible is a key aim for PLA. de Laat & Schreurs (2013) demonstrated that LA techniques can visualize informal professional networks. Using a tool based on social network analysis – the network awareness tool (NAT) – they detected multiple isolated networks of teachers within a single organization. More recent work uses wearable devices to track professional networks in health care settings (Endedijk, 2016). By exploiting these informal professional networks, organizations have a mechanism to improve human social capital and learning. However, the study illustrates the limited extent of connections across professional networks, so multiple methods of exploiting networks are needed.

Learning Layers[3] exploits networks and relationships at the individual, organizational, and inter-organizational levels to improve performance. Professionals working within and across regional small to medium enterprises (SMEs) in different European countries work together in web-based, networked environments to build ideas and knowledge. The system exploits semantic technologies to analyze and support the co-production of knowledge by connecting people and making recommendations (Ley et al., 2014).

Technology environments tested in the health and construction sectors illustrate ways professionals work and learn together at three levels. As people work together, they learn how to develop different types of knowledge including technical knowledge (know-what), procedural or practical knowledge (know-how), and scientific or theoretical knowledge (know-why) to help them make informed choices as to how they will carry out new work tasks (Attwell et al., 2013). At the organizational level, as people collaborate across different SMEs, the individual organizations share knowledge and learn. Thirdly, the companies are grouped in national clusters, so learning occurs across organizations (Dennerlein et al., 2015). LA tools are being co-developed with the professionals themselves and with key stakeholders. An open design library[4] is used to store and disseminate ideas for professional learning while an open developer library[5] hosts prototype tools and codes for developers to work on. The principle that professionals themselves have the best knowledge about learning for highly specialized roles is central to several LA systems and tools.

Making use of Specialist Expertise

Responsive Open Learning Environments (ROLE)[6] are being developed to enable professionals to adapt to and deal with change and uncertainty in their work and learning (Kirschenmann, Scheffel, Friedrich, Niemann, & Wolpers, 2010). Rather than using an all-encompassing, enterprise system, the learning environment can be personalized by each individual learner. Each professional can browse and select a set of web-based software tools with specific functions that support learning for a specific role. The tools can be combined to form new components and functionalities. By establishing a unique combination of tools and resources, the professional embeds her own expertise within the environment. This combination of tools and resources can be reproduced and adapted to support other professionals with similar work and learning needs.

[3] learning-layers.eu
[4] odl.learning-layers.eu
[5] developer.learning-layers.eu
[6] role-project.eu

The basic principle that underpins ROLE is well defined; in highly specialized roles, professionals themselves are best placed to decide on their learning needs (Kroop, Mikroyannidis, & Wolpers, 2015). However, from a professional learning perspective, there are two main overarching difficulties. First, professionals may not have the skills to implement their decisions. ROLE is based on an open framework that allows the development of technology tools or widgets that support specific aspects of professional learning. These widgets are developed through open competitions, where developers are encouraged to create new tools and functionalities. Ideally, developers who have a deep understanding of the job role would create the widgets: if the professionals themselves can write the algorithms for the widgets, then their expertise is embedded within the code. However, not all professionals can write code, so, for many professions the widgets tend to be developed by people who don't have a deep understanding of the job role. The second difficulty is that professionals have to be able to identify and act upon their learning needs. The ROLE environment includes a course on becoming a self-regulated learner.[7] However, a problem is that while some aspects of self-regulation can be learned, such as developing strategies to set learning goals, other facets of self-regulation, such as self-confidence, are developed through lifelong experiences.

Encouraging Active Learning

Two other examples of LA systems based on self-regulated learning theory are *LearnB* and *Mirror. LearnB* has been piloted in the automotive industry (Siadaty, Gašević & Hatala, 2016). What this tool has in common with the previous systems is that it encourages professionals to self-regulate their learning. The tool is designed around a self-regulated learning framework, SRL@Work, which is used to gather data on factors that influence self-regulation. These factors include how learning goals are planned and the specific range of activities that people engage in as they share and build knowledge for work. Learn B uses Social Semantic Web technologies to gather and analyze these data in order to identify and connect people with similar learning and development goals (Siadaty et al., 2012). Common goals are identified and analyzed using the semantic capabilities of the system. Then the system uses social technologies to recommend topics people might benefit from learning, based on the learning patterns of others.

The LearnB system serves as a "developmental radar" allowing professionals to source and assess potentially useful connections with other people and with relevant knowledge (Siadaty, Jovanović, Gašević, Jeremić, & Holocher-Ertl, 2010; Holocher-Ertl et al., 2012). It can be used to advise professionals on their learning strategies while monitoring their learning progress. The idea here is that people might learn effectively by using strategies that have been effective for other people with analogous experience. The system supports professionals not only in documenting their learning experiences, but also in making these experiences available for others who might benefit from learning in a similar way in the future. By documenting learning experiences, it is possible to share and compare with the performance of others or against organizational benchmarks. It might be useful, for example, to know that it takes an average of six months' experience to become competent in a new procedure. On the other hand, it may be reassuring to know that a new skill can be learned in a few hours (Siadaty et al., 2012).

The Learn B trial demonstrated the importance of integrating the system with active development of professionals' self-regulated learning skills. Professionals who used Learn B benefitted from having their attention directed towards useful knowledge, sometimes sourced from contexts or departments different from those in which they worked. They also perceived the usefulness of having access to data on other peoples' informal work and learning practices that helped them to understand the ways other people learned. They perceived that they benefitted from updates about their social context – knowing, for example, the actions other people were taking to learn – and by being informed about how the available learning resources were used by their colleagues (Siadaty, 2013). However, a critical point is that these professionals were operating within a traditional organizational culture with the sort of "top down" competence systems discussed by Clow (2013) where the organization predetermines the competencies needed for each job role and recommends the ways people demonstrate how they learn these capabilities.

Mirror[8] is an analytics-based system that supports professionals in learning from their own and others experiences. Reflection is a significant component of self-regulated learning that may improve learning and performance through motivational and affective factors (Littlejohn & Hood, 2016). The Mirror system is based on a set of applications ("Mirror" apps) designed to facilitate informal learning during work (Kump, Knipfer, Pammer, Schmidt, & Maier, 2011). These apps were used in case worker settings to support analysis of individual and team actions. These reflections allowed both individuals and teams to learn which practices had the most impact within their organization. Eval-

[7] ROLE Course in Becoming a Self-regulated Learner (http://www.open.edu/openlearnworks/course/view.php?id=1490

[8] www.mirror-project.eu

uation studies found a clear link between individual and team learning and organizational learning (linked to Human Resource procedures, rewards, and promotions) (Knipfer, Kump, Wessel, & Cress, 2013). Without a parallel shift in the culture and the mindsets of people within the organization, new analytics systems will have limited impact.

CONCLUSION: FUTURE PROFESSIONAL LEARNING ANALYTICS

Novel approaches to LA are already supporting professionals in improving their performance. Analysis of these approaches point to emerging themes that can inform future work.

Several approaches to analytics use machine-based analytics to augment human intelligence. However, the connection between the system and the human is a point of risk for a number of reasons. First, professionals must be able to identify and act upon their learning needs, so the ability to self-regulate learning is critical to the success of many analytics techniques. Second, without a parallel shift in the culture and the mindsets of people within the organization, learning systems based on analytics will have limited impact. Implementation of approaches to LA should consider these human elements.

Some LA techniques use network analysis to gather data. Workplace learning is most effective when learning processes are aligned with organizational and project goals (Kimmerle, Cress & Held, 2010; Kozlowski & Klein, 2000). Network analysis should aim to align individual learning activities intelligently with organizational learning goals.

Other promising approaches to PLA are based on the development of software applications (or apps). Ideally, professionals who have the specific expertise would write the code. However, not all professionals have the ability write code, so, for many professions apps tend to be developed by people who do not have a deep understanding of the job role. This problem is likely to be more significant in non-technical sectors.

Many PLA methods aim to embed agile approaches to learning through capturing and disseminating expertise. There are a number of problems associated with this approach; for example, not all expertise can be codified – expertise development pathways are difficult to model since each individual has a unique baseline of prior knowledge and professionals to be actively involved in learning new expertise. These difficulties illustrate that the analytics solutions have to not only address the technical and practical aspects of expertise development, but deal with affective and motivational issues as well.

Although in its infancy, professional learning analytics is set to form a foundation for future learning and work. However, careful attention has to be paid to the alignment of the knowledge on how professionals learn with analytics applications.

REFERENCES

Attwell, G., Heinemann, L., Deitmer, L., Kämaäinen, P. (2013). Developing PLEs to support work practice based learning. In eLearning Papers #35, Nov. 2013. http://OpenEducationEuropa.eu/en/elearning_papers

Argyris, C., & Schon, D. A. (1974). *Theory in practice: Increasing professional effectiveness*. San Francisco, CA: Jossey-Bass, 1974.

Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education, 26*(2), 600–614.

Berendt, B., Vuorikari, R., Littlejohn, A., & Margaryan, A. (2014). Learning analytics and their application in technology-enhanced professional learning. In A. Littlejohn & A. Margaryan (Eds.), *Technology-enhanced professional learning: Processes, practices and tools* (pp. 144–157). London: Routledge.

Boud, D., & Garrick, J. (1999). *Understanding learning at work*. Taylor & Francis US.

Billett, S. (2004). Workplace participatory practices: Conceptualising workplaces as learning environments. *Journal of Workplace Learning, 16*(6), 312–324.

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 92–101). New York: ACM.

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology & Society*, 15(3), 3–26.

Clow, J. (2013) Work practices to support continuous organisational learning. In A. Littlejohn & A. Margaryan (Eds.), *Technology-enhanced professional learning: Processes, practices, and tools* (pp. 17–27). London: Routledge.

Collin, K. (2008). Development engineers' work and learning as shared practice. *International Journal of Lifelong Education*, 27(4), 379–397.

Convertino, G., Grasso, A., DiMicco, J., De Michelis, G., & Chi, E. H. (2010). Collective intelligence in organizations: Towards a research agenda. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work* (CSCW 2010), 6–10 February 2010, Atlanta, GA, USA (pp. 613–614). New York: ACM. http://research.microsoft.com/en-us/um/redmond/groups/connect/CSCW_10/docs/p613.pdf

Dennerlein, S., Theiler, D., Marton, P., Santos, P., Cook, J., Lindstaedt, S., & Lex, E. (2015) KnowBrain: An online social knowledge repository for informal workplace learning. *Proceedings of the 10th European Conference on Technology Enhanced Learning* (EC-TEL '15), 15–17 September 2015, Toledo, Spain. http://eprints.uwe.ac.uk/26226

Endedijk, M. (2016) Presentation at the Technology-enhanced Professional Learning (TePL) Special Interest Group, Glasgow Caledonian University, May 2016.

Eraut, M. (2000). Non-formal learning and tacit knowledge in professional work. *British Journal of Educational Psychology*, 70(1), 113–136.

Eraut, M. (2004). Informal learning in the workplace. *Studies in Continuing Education*, 26(2), 247–273.

Eraut, M. (2007). Learning from other people in the workplace. *Oxford Review of Education*, 33(4), 403–422.

Engeström, Y. (1999). Innovative learning in work teams: Analyzing cycles of knowledge creation in practice. *Perspectives on activity theory*, 377-404.

Engeström, Y. (2008). *From teams to knots: Activity-theoretical studies of collaboration and learning at work*. Cambridge, UK: Cambridge University Press.

Ferguson, R. (2012) Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5–6), 304–317.

Fuller, A., Ashton, D., Felstead, A., Unwin, L., Walters, S., & Quinn, M. (2003). The impact of informal learning for work on business productivity. London: Department of Trade and Industry.

Fuller, A., & Unwin, L. (2004). lntegrating organizational and personal development. *Workplace learning in context*, 126.

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education*, 23, 18–26.

Harteis, C., & Billett, S. (2008). The workplace as learning environment: Introduction. *International Journal of Educational Research*, 47(4), 209–212.

HEC Report (2016) From bricks to clicks: Higher Education Commission Report http://www.policyconnect.org.uk/hec/sites/site_hec/files/report/419/fieldreportdownload/frombrickstoclicks-hecreportforweb.pdf

Holocher-Ertl, T., Fabian, C. M., Siadaty, M., Jovanović, J., Pata, K., & Gašević, D. (2011). Self-regulated learners and collaboration: How innovative tools can address the motivation to learn at the workplace? *Proceedings of the 6th European Conference on Technology Enhanced Learning: Towards Ubiquitous Learning* (EC-TEL 2011), 20–23 September 2011, Palermo, Italy (pp. 506–511). Springer Berlin Heidelberg.

IBM (2008) Making change work. IBM Global Study. http://www-07.ibm.com/au/pdf/making_change_work.pdf

de Laat, M., & Schreurs, B. (2013). Visualizing informal professional development networks: Building a case for learning analytics in the workplace. *American Behavioral Scientist*, 57(10), 1421–1438.

Kimmerle, J., Cress, U., & Held, C. (2010). The interplay between individual and collective knowledge: Technologies for organisational learning and knowledge building. *Knowledge Management Research & Practice*, 8(1), 33–44.

Kirschenmann, U., Scheffel, M., Friedrich, M., Niemann, K., & Wolpers, M. (2010). Demands of modern PLEs and the ROLE approach. In M. Wolpers, P. A. Kirschner, M. Scheffel, S. N. Lindstaedt, & V. Dimitrova (Eds.), *Proceedings of the 5th European Conference on Technology Enhanced Learning, Sustaining TEL: From Innovation to Learning and Practice* (EC-TEL 2010), 28 September–1 October 2010, Barcelona, Spain (pp. 167–182). Springer Berlin Heidelberg.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory, research and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). San Francisco, CA: Jossey-Bass.

Knipfer, K., Kump, B., Wessel, D., & Cress, U. (2013). Reflection as a catalyst for organisational learning. *Studies in Continuing Education*, 35(1), 30–48.

Kroop, S., Mikroyannidis, A., & Wolpers, M. (Eds.). (2015). *Responsive open learning environments: Outcomes of research from the ROLE project*. Springer.

Kump, B., Knipfer, K., Pammer, V., Schmidt, A., & Maier, R. (2011) The role of reflection in maturing organizational know-how. *Proceedings of the 1st European Workshop on Awareness and Reflection in Learning Networks* (@ EC-TEL 2011), 21 September 2011, Palermo, Italy.

Ley, T., Cook, J., Dennerlein, S., Kravcik, M., Kunzmann, C., Pata, K., & Trattner, C. (2014). Scaling informal learning at the workplace: A model and four designs from a large-scale design-based research effort. *British Journal of Educational Technology*, 45(6), 1036–1048.

Littlejohn, A., & Hood, N. (2016, March 18). How educators build knowledge and expand their practice: The case of open education resources. *British Journal of Educational Technology*. doi:10.1111/bjet.12438

Littlejohn, A., & Margaryan, A. (2013). Technology-enhanced professional learning: Mapping out a new domain. In A. Littlejohn & A. Margaryan (Eds.), *Technology-enhanced professional learning: Processes, practices, and tools* (Chapter 1). Taylor & Francis.

Littlejohn, A., Milligan, C., & Margaryan, A. (2012). Charting collective knowledge: Supporting self-regulated learning in the workplace. *Journal of Workplace Learning*, 24(3), 226–238.

Milligan, C., Littlejohn, A., & Margaryan, A. (2014). Workplace learning in informal networks. *Reusing Open Resources: Learning in Open Networks for Work, Life and Education*, 93.

Nistor, N., Derntl, M., & Klamma, R. (2015). Learning analytics: Trends and issues of the empirical research of the years 2011–2014. *Proceedings of the 10th European Conference on Technology Enhanced Learning: Design for Teaching and Learning in a Networked World* (EC-TEL 2015), 15–17 September 2015, Toledo, Spain (pp. pp. 453–459). Springer.

Rienties, B., Boroowa, A., Cross, S., Kubiak, C., Mayles, K., & Murphy, S. (2016). Analytics4Action evaluation framework: A review of evidence-based learning analytics interventions at the Open University UK. *Journal of Interactive Media in Education*, 2016(1). www-jime.open.ac.uk/articles/10.5334/jime.394/galley/596/download/

Siadaty, M. (2013). *Semantic web-enabled interventions to support workplace learning*. Doctoral dissertation, Simon Fraser University, Burnaby, BC. http://summit.sfu.ca/item/12795

Siadaty, M., Gašević, D., & Hatala, M. (2016). Associations between technological scaffolding and micro-level processes of self-regulated learning: A workplace study. *Computers in Human Behavior*, 55, 1007–1019.

Siadaty, M., Gašević, D., Jovanović, J., Pata, K., Milikić, N., Holocher-Ertl, T., ... & Hatala, M. (2012). Self-regulated workplace learning: A pedagogical framework and semantic web-based environment. *Educational Technology & Society*, 15(4), 75–88.

Siadaty, M., Jovanović, J., Gašević, D., Jeremić, Z., & Holocher-Ertl, T. (2010). Leveraging semantic technologies for harmonization of individual and organizational learning. In M. Wolpers, P. A. Kirschner, M. Scheffel, S. N. Lindstaedt, & V. Dimitrova (Eds.), *Proceedings of the 5th European Conference on Technology Enhanced Learning, Sustaining TEL: From Innovation to Learning and Practice* (EC-TEL 2010), 28 September–1 October 2010, Barcelona, Spain (pp. 340–356). Springer Berlin Heidelberg.

Siemens, G., & Long, P. (2011) Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 31–40. http://net.educause.edu/ir/library/pdf/ERM1151.pdf

Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review*, 3(2), 130–154.

Tynjälä, P., & Gijbels, D. (2012). Changing world: Changing pedagogy. In P. Tynjälä, M.-L. Stenström, & M. Saarnivaara (Eds.), *Transitions and transformations in learning and education* (pp. 205–222). Springer Netherlands.

Wen, M., Yang, D., & Rosé, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper et al. (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July 2014, London, UK. International Educational Data Mining Society. http://www.cs.cmu.edu/~mwen/papers/edm2014-camera-ready.pdf

Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013, April). Improving retention: Predicting at-risk students by analysing clicking behaviour in a virtual learning environment. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 145–149). New York: ACM.

SECTION 4

INSTITUTIONAL STRATEGIES & SYSTEMS PERSPECTIVES

Chapter 24: Addressing the Challenges of Institutional Adoption

Cassandra Colvin[1], Shane Dawson[1], Alexandra Wade[2], Dragan Gašević[3]

[1]Teaching Innovation Unit, University of South Australia, Australia
[2]School of Health Sciences, University of South Australia, Australia
[3]Schools of Education and Informatics, The University of Edinburgh, United Kingdom
DOI: 10.18608/hla17.024

ABSTRACT

Despite increased funding opportunities, research, and institutional investment, there remains a paucity of realized large-scale implementations of learning analytics strategies and activities in higher education. The lack of institutional exemplars denies the sector broad and nuanced understanding of the affordances and constraints of learning analytics implementations over time. This chapter explores the various models informing the adoption of large-scale learning analytics projects. In so doing, it highlights the limitations of current work and proposes a more empirically driven approach to identify the complex and interwoven dimensions impacting learning analytics adoption at scale.

Keywords: Learning analytics adoption, learning analytics uptake, leadership

The importance of data and analytics for learning and teaching practice is strongly argued in the education policy and research literature (Daniel, 2015; Siemens, Dawson, & Lynch, 2013). The insights into teaching, learning, student-experience, and management activities that learning analytics afford are touted to be unprecedented in scale, sophistication, and impact (Baker & Inventado, 2014). Not only do learning analytics have the capacity to provide rich understanding of practices and activities occurring within institutions, they also have the potential to mediate and shape future activity through, for example, predictive modelling, personalization of learning, and recommendation systems (Conde & Hernández-García, 2015).

Despite increased funding opportunities, research, and institutional investment, there remains a paucity of realized large-scale implementations of learning analytics strategies and activities in higher education (Ferguson et al., 2015), thus denying the sector broad and nuanced understanding of the affordances and constraints of learning analytics implementations over time. Part of the explanation for the lack of enterprise exemplars may lie in the relative nascency of learning analytics as a discipline and a perceived lack of time for learning analytics programs and implementations to fully develop and mature. However, this explanation does not adequately capture the complexity of issues mediating systemic uptake of learning analytics (Arnold, Lynch, et al., 2014; Ferguson et al., 2015; Macfadyen, Dawson, Pardo, & Gašević, 2014). Although learning analytics is relatively new to higher education, we suggest there have been sufficient investments made in time and resources to realize the affordances such activities can bring to education at a whole-of-institution scale. Indeed, a small number of institutions have been able to implement large-scale learning analytics programs with demonstrable impact on their teaching and learning outcomes (Ferguson et al., 2015). However, these examples remain the exception in a sector where, for a large number of institutions, organizational adoption of learning analytics either remains a conceptual, unrealized aspiration or, where operationalized, is often narrow and limited in scope and impact (Ferguson et al., 2015).

A burgeoning body of conceptual literature has recently begun to explore this vexing issue (Arnold, Lonn, & Pistilli, 2014; Arnold, Lynch, et al., 2014; Ferguson et al., 2015; Macfadyen et al., 2014; Norris, Baer, Leonard, Pugliese, & Lefrere, 2008). This literature proffers multiple frameworks intended to capture, and elicit insight into, dimensions and processes mediating learning analytics adoption. In addition to aiding conceptual understanding, this literature also has a heuristic value, guiding institutions through implementation stages and considerations. Given the limited empirical research exploring learning analytics deployment (Ferguson et al., 2015), it is probable that many managers turn to this small body of conceptual literature for inspiration and insight when planning and administering learning analytics initiatives. The present chapter reviews this body of literature to glean from it not only insight into what it identifies as dimensions and processes important for effective institutional implementations of learning analytics, but also to gauge the merit of the models as guides for institutional managers. We then compare and contrast the findings of this review of the literature with those from a recent study that examined actual learning analytics implementations across a large cohort of Australian universities to proffer empirical understanding into the processes and factors affording them (Colvin et al., 2015).

REVIEW OF CURRENT MODELS OF LEARNING ANALYTICS DEPLOYMENT

Review of extant learning analytics implementation models and frameworks revealed three primary groups of literature: 1) those focused on the antecedents to learning analytics outcomes (learning analytics inputs models); 2) those focused on the outcomes of learning analytics (learning analytics outputs models); and 3) process models that sequentially map and operationalize tasks underpinning learning analytics implementations. An overview of these different models, and their conceptual and empirical contribution to understanding factors shaping institutional learning analytics implementations, follows.

Learning Analytics Inputs Models

Frameworks in this body of literature tend to present learning analytics implementations as a consequence of antecedent affordances incorporating dimensions such as leadership, governance, technology, capacity, and culture. Notable in this literature is the US-based EDUCAUSE Centre for Analysis and Research (ECAR; see ECAR, 2015) Analytics Maturity Index for Higher Education (Bichsel, 2012). Their model, informed by data elicited through surveys and focus group interviews with industry professionals, operationalizes learning analytics implementations across six dimensions of activity including culture, process, data/reporting/tools, investment, expertise, and governance/infrastructure. Each input dimension is scaffolded across a continuum of five maturity levels designed to assist institutions in determining their level of progress within each level. The criticality of each input dimension for a successful learning analytics implementation is assumed.

Similar to the ECAR model is the Learning Analytics Readiness Instrument (LARI) (Arnold, Lonn, & Pistilli, 2014; Oster, Lonn, Pistilli, & Brown, 2016), a tool designed to assist institutions in assessing their level of "readiness" for analytics implementations. The original version of the instrument (Arnold, Lonn, & Pistilli, 2014) identified five dimensions – 1) ability, 2) data, 3) culture and process, 4) governance and infrastructure, and overall 5) readiness perceptions – as essential for achieving "the optimal environment for learning analytics success" (p. 2), although it is unclear how the five elements were initially determined. A more recent factor analysis of survey data from 560 participants across 24 institutions was used to refine the LARI. The five dimensions were slightly altered, and their relative salience was revealed. However, salience was measured according to participant perception, and not against learning analytics implementations outcomes.

The Organizational Capacity Analytics Framework (Norris & Baer, 2013) is also founded on insight gleaned from learning analytics specialists as to dimensions they consider important in shaping analytics adoption. The authors interviewed managers from 40 institutions in the United States, and data collected through these interviews led to the generation of five dimensions deemed to be critical organizational capacity factors. These dimensions were 1) technology infrastructure, 2) processes and practice, 3) culture and behaviours, 4) skills and values, and 5) leadership. Notable in their Organizational Capacity for Analytics Framework is the presentation of the dimensions as interconnected and overlapping, thereby highlighting their interdependent nature (Norris & Baer, 2013, p. 31). While the framework operationalizes three maturity levels for each of the dimensions, it does not examine their relative salience.

Finally, Drachsler and Greller's model (referred to by the authors as an ontology; Drachsler & Greller, 2012; Greller & Drachsler, 2012) also captures the interdependent and recursive nature of dimensions mediating learning analytics implementations. General morphological analysis (cf. Ritchey, 2011, in Greller & Drachsler, 2012) was applied to data solicited from media scanning, interviews with senior experts, and a cognitive mapping exercise. This model identifies six core activity areas as "critical" to "ensure an appropriate exploitation of learning analytics" (Drachsler & Greller, 2012, p. 120): 1) competences, 2) constraints (privacy/ethics), 3) technologies, 4) education data, 5) objectives, and 6) stakeholders. While Drachsler and Greller (2012) deem each of these six dimensions to be "critical," they observe that their salience is not uniform, noting "some dimensions are vaguer than others" (p. 44).

Common to many input models is the conceptualization of learning analytics implementations as non-linear, emergent from, and afforded by, the interplay of multiple, interconnected input dimensions. While often informed by opinion solicited through focus group and survey methods (Drachsler & Greller, 2012; Greller & Drachsler, 2012; Norris & Baer, 2013; Oster et al., 2016), these models are essentially conceptual. They identify dimensions that institutional representatives perceive must be considered to mount an effective learning analytics implementation, yet they do not empirically interrogate the dimensions against actual learning analytics implementations. Similarly, while the input models suggest interrelationships between antecedent dimensions, the nature of these interrelationships, and their impacts on learning analytic implementation outputs, are also not empirically explored. Therefore, while the models offer leaders charged with the task of implementing learning analytics programs in their institutions insight into the antecedent dimensions necessary to effect an implementation, they present little guidance on how such implementations could look in action. Further, the relative salience of each dimension within the models is underexplored, limiting insight into how institutions might best prioritize actions and resources (although a limited number of models do accommodate gradations of maturity within each dimension; Arnold, Lonn, & Pistilli, 2014; Bichsel, 2012; Norris & Baer, 2013).

Learning Analytics Outputs Models

This second body of learning analytics models and frameworks defines and represents learning analytics implementations as a linear process, unfolding over time, and involving different levels of readiness and maturity. An early model in this literature that still appears to have resonance in the sector is Davenport and Harris's (2007) Analytics Framework, which conceptualizes analytics as a maturing process from query and reporting applications through to formal analytics functions such as forecasting and predictive modelling. Siemens, Dawson, and Lynch's (2013) Learning Analytics Sophistication Model integrates analytic capability and systems deployment along a continuum of increasing maturity. Five key stages of maturity are identified, each of these further operationalized into sample exemplars. For instance, an early stage deployment would feature basic reports and log data whereas a mature deployment would feature predictive models and personalized learning.

A primary benefit of outputs models is that they provide a means for institutions to objectively assess the maturity (or capacity) of their activities and processes against a matrix of desired outcomes. However, many outputs models are limited in scope and typically predicated on a uni- or bi-dimensional conceptual lens (such as the sophistication of analytic techniques employed). While outputs models advocate a vision of learning analytics implementations outcomes, they often fail to identify or critically examine all of the dimensions or mechanisms needed to generate the LA implementation outcomes they in fact advocate. Finally, a risk of many outputs models is that they conceptualize progression as a linear and hierarchical process, culminating in an "essentialized," perhaps even "utopian" vision of learning analytics, one that is typically conceptual, removed from context, possibly predicated on an assumed universality, and not necessarily capturing what might be possible or desirable within the scope of a particular institution's operating context.

Process Models

This third body of literature (Foreman, 2013a, 2013b; Norris & Baer, 2013) sequentially maps key processes or "steps" underpinning learning analytics implementations. It focuses on the how of implementing a learning analytics program, rather than what the outcomes should look like (outputs model) or involve (inputs model). Process models are both linear (Foreman, 2013a, 2013b) or circuitous (Norris and Baer's Action Plan for Analytics, 2013) and are typically focused on specific elements within a broader learning analytics implementation (for instance, the implementation of a learning management system (LMS; Foreman, 2013a, 2013b), or strategy development (cf. Norris and Baer's Action Plan for Analytics, 2013). However, emerging literature (Ferguson et al., 2015) presents processual models that better reflect the breadth and complexity of learning analytics implementations, arguing that the insight this conceptualization affords is critical for institutions wishing to apply learning analytics "at scale." Most notable is Ferguson and colleagues' (2015) advocacy for adopting the RAPID Outcomes Mapping Approach (ROMA) for a learning analytics implementation. This model presents inputs dimensions in an operational sequence involving seven key steps from formulation of initial objectives through to final evaluation. However, Ferguson's model, in essence, is conceptually generated. While there is little evidence of the model's empirical validation, it has been applied as a lens to describe learning analytics implementations at universities in Australia and the UK, argued by the authors to demonstrate the model's potential guide and give "confidence" to institutions (Ferguson et al., 2015).

WHAT DO THE MODELS TELL US?

While reviewing these models affords insight into dimensions and processes that mediate learning analytics deployments, it also reveals the models' conceptual and operational limitations. These include their adoption of a limited, or unidimensional lens to scrutinize complex, multidimensional phenomena; their inability to integrate antecedent dimensions and outcomes in the one model; and their limited insight into the relative salience or criticality of each of the identified mediating dimensions. Simply put, while the models afford insight, they do not fully capture the breadth of factors that shape LA implementations, thus curtailing their ability to present managers with the nuanced, situated, fine-grained insight they require to guide them through learning analytics implementations.

Notwithstanding, a number of mediating dimensions, or elements, were found to be common to most models, suggesting them to be particularly salient. These included technological readiness, leadership, organizational culture, staff and institutional capacity, and strategy. Discussion surrounding how these dimensions are operationalized in the models follows.

Technological Readiness

As learning analytics is essentially grounded in the affordance of technology to offer access and insight into electronic data, it is not surprising that technology features in LA implementation literature as a "foundational element" (Arnold, Lynch, et al., 2014, p. 258; refer also to Greller & Drachsler, 2012; Siemens & Long, 2011). However, operationalizations of dimensions described as technological readiness vary across the models. For instance, some models emphasize the need for a robust technology infrastructure that can collect, store, and transform data (Arnold, Lynch, et al., 2014, p. 258), while others reinforce the need for integrated systems (Dawson, Heathcote, & Poole, 2010; Siemens & Long, 2011), appropriate analytics tools (Norris & Baer, 2013), and security and privacy controls and processes (for instance, the ECAR Analytics Maturity Index for Higher Education Model in Bichsel, 2012). Empirically, the potentially militating role of technology as a constraining element in learning analytics implementations was noted in studies undertaken by Dawson and colleagues (Dawson et al., 2010; Macfadyen & Dawson, 2012).

Leadership

The criticality of leadership for sustainable implementations of learning analytics at scale is well recognized conceptually (Arnold, Lonn, & Pistilli, 2014; Arnold, Lynch, et al., 2014; Laferrière, Hamel, & Searson, 2013; Norris & Baer, 2013; Siemens et al., 2013) and empirically (Graham, Woodfield, & Harrison, 2013; Norris & Baer, 2013). This literature advocates the importance of "committed" and "informed" leadership ground in a "deep scholarly understanding" of learning analytics to facilitate uptake and integration (Arnold, Lynch, et al., 2014, p. 260). While there is an obvious need for committed senior leadership, particularly in projects of scale and complexity (Norris & Baer, 2013), there is a lack of consensus and commentary regarding how such leadership is conceptualized. For example, Laferrière et al. (2013) operationalized leadership through a uni- or limited dimensionality lens. In contrast, Arnold, Lynch et al. (2014) recognize leadership as a multilayered, multidimensional phenomenon. Leadership is also operationalized in the literature as leadership style (Owston, 2013), or leadership behaviour and influence (Laferrière et al., 2013), while other literature refers to leadership's structure (Accard, 2015; Carbonell, Dailey-Hebert, & Gijselaers, 2013) and strength (cf. Kotter & Schlesinger, 2008, in Arnold, Lynch, et al., 2014). Gaining particular traction is research advocating complexity (Hazy & Uhl-Bien, 2014) or distributed leadership models (Bolden, 2011) to aid analytics implementation and uptake.

Organizational Culture

Organizational culture, defined as an institution's "norms, beliefs and values" (Carbonell et al., 2013, p. 30), has also been identified as a key mediator of learning analytics implementations (Arnold, Lonn, & Pistilli, 2014; Carbonell et al., 2013; Greller & Drachsler, 2012; Macfadyen & Dawson, 2012). Prominent in this literature is an emphasis on staff "awareness and acceptance of data" (Arnold, Lonn, & Pistilli, 2014, p. 164), a recognition of the potentially militating influence of an institution's "historical pedagogical [and] socio-cultural assumptions" vis-à-vis educational practice (Arnold, Lynch, et al., 2014, p. 259), organizational "routines" (Carbonell et al., 2013, p. 29), and even staff anxiety regarding organizational, pedagogical, and IT education change (Houchin & MacLean, 2005). Empirical insight into the impact of an insufficiently prepared and receptive organizational culture has been offered in Macfadyen and Dawson's (2012) research into a failed implementation at a large Canadian university. The researchers observed that the institution's failure to generate a shared, willing, and receptive appreciation of learning analytics potential was a key reason for the organization's failure to roll out a coherent and successful learning analytics strategy.

Staff and Institutional Capacity

"Optimal" (Greller & Drachsler, 2012, p. 51) learning analytics outcomes are contingent on the ability of staff to effectively analyse, interpret, and meaningfully respond to analytics intelligence (Bichsel, 2012; Norris & Baer, 2013). However, it cannot be assumed that stakeholders possess the necessary analytical or interpretive data skills demanded of learning analytics. Norris and Baer (2013) observe that "many institutional leaders overestimate their enterprise's capacity in data, information, and analytics" (p. 40). Drachsler and Greller's (2012) research into the dimensions required in learning analytics implementations distinguishes between hard and soft dimensions: soft dimensions refer to the "human factors" that shape learning analytics effectiveness, notably "competences and acceptance" (p. 43); "hard" dimensions refer to nonhuman, less subjective elements, including technology, data, and algorithms. Distinguishing between hard and soft dimensions highlights the need for institutions to consider learning analytics implementations as extending beyond technical, infrastructure issues, to include sociocultural concerns. The successful adoption of learning analytics requires capacity building across these two domains.

The Learning Analytics Readiness Instrument (LARI; Arnold, Lonn, & Pistilli, 2014) introduces a different conceptualization of capacity: institutional readiness – that is, a measure of how "ready" an institution is to implement a learning analytics initiative. They operationalize readiness across five dimensions: 1) ability, 2) data, 3) culture and process, 4) governance and infrastructure, and 5) overall readiness perception. Their conceptualization highlights the multilayered nature of capacity, noting that it presents at macro (i.e., broad, whole of institution) and micro (the level of the individual stakeholder) levels. Finally, in contrast with the focus on technical, critical, and interpretative capacity, Siemens, Dawson, and Lynch (2013) remind us of the relationship between learning analytics and teaching and learning practice, suggesting that capacity should also encompass the ability of staff to effectively "link" pedagogy and analytics.

Strategy

Conceptual literature advocates the development of a clear vision and purpose of learning analytics through the development of policy and procedures. For example, Arnold, Lonn, and Pistilli's (2014) conceptually developed Learning Analytics Readiness Instrument (LARI) subsumes policy into a broader category of mediating dimensions labelled governance and infrastructure. Further, Norris and Baer's Organizational Capacity for Analytics model (2013) identifies "Processes and Practices" as one of five key mediating dimensions of organizational capacity for LA, operationalizing this as "routinized processes and workflows to leverage […] analytics, actions, and interventions" (p. 31). However, and by contrast, empirical studies stress the importance of strategy setting, and emphasize the need for aligned policies and objectives (Macfadyen & Dawson, 2012; Owston, 2013), Macfadyen and Dawson (2012) declaring, in their analysis of a failed learning analytics program, that the establishment of an organizational strategy and vision was "critical" (p. 150) for learning analytics implementations.

In Sum

This brief literature review found that learning analytics models can provide managers with valuable insight into processes and dimensions shaping learning analytics implementations. Specifically, five dimensions were frequently highlighted across multiple frameworks as having impact on learning analytics implementation outcomes: technological readiness, leadership, organizational culture, staff and institutional capacity for learning analytics, and learning analytics strategy. However, as we have noted, operationalizations of these dimensions varied across the literature. Furthermore, the models afforded little insight into the relative salience, or criticality, of the dimensions. We suggest that the differing conceptualizations and operationalizations of the dimensions referred to in the literature have the potential to mediate how institutions engage with and interpret the many learning analytics frameworks available to them.

Further, and significantly, the literature introduced in this review is predominantly conceptual. We argue that the lack of empirical research into learning analytics implementations has hindered our understanding of the processes and dimensions that mediate them. While conceptual literature affords insight, it risks presenting an idealized model of learning analytics that might not adequately capture its full complexity and nuance. Where empirical techniques have been employed (such as soliciting data through surveys and focus groups), there is little detail surrounding construct validity. Accordingly, relationships between the different dimensions in the models appear to be largely untested. As observed earlier in this chapter, the relative immaturity of learning analytics programs in higher education institutions contributes in part to this empirical paucity surrounding learning analytics. However, we argue that the burgeoning, albeit nascent implementations found across higher education institutions provide an opportunity to empirically scrutinize how learning analytics implementations are currently being performed and mediated in context.

BRINGING IN THE EMPIRICAL

Recent research based in Australia has sought to address this research gap. Colvin et al. (2015) undertook a large national study investigating learning analytics implementations across the Australian higher education sector. Data were solicited through qualitative interviews with senior leaders charged with responsibility for implementing learning analytics at 32 universities. Utilizing a mixed-method methodology, the study identified six dimensions (inductively generated) that had a statistically significant impact on learning analytics implementations (out of the 27 dimensions identified in the data). Largely reflective of prior literature, four of the dimensions found to have impact included effective and distributed *stakeholder* engagement, *technological* capacity, clear vision and *strategy*, and influential *leadership*. Two other dimensions were revealed, namely institutional context (including an institution's student and institutional profile) and conceptualization (how an institution constructed and understood learning analytics). The former of these dimensions, institutional context, reminds us that learning analytics is situated in practice, shaped by an array of social and institutional structural elements unique to each institution's context. By contrast, the "conceptualization" dimension related to an institution's underlying epistemological and ontological position vis-à-vis learning analytics implementations. While institutions were found to have diverging understandings, aspirations, and visions of learning analytics, relationships were found between how learning analytics was conceptualized by an institution and how it was actually implemented. Simply, the findings of this study suggest that how learning analytics is understood, and the meaning assigned to it, appears to shape how it is implemented. Further, and significantly, cluster analysis performed in the research suggested the emergence of two trajectories of learning analytics implementation in the Australian higher education context. Within each of the clusters, there was congruence in how the conceptualization, readiness (antecedent), and implementation dimensions were performed and experienced across institutions. One cluster appeared to privilege a more instrumental conceptualization and a retention, student at-risk operationalization of learning analytics. By contrast, institutions in the second cluster were also invested in retention-focused learning analytics activity, but supplemented this with activity aimed to elicit insight into, and inform teaching and learning. Colvin at al.'s (2015) finding that there appear to be two diverging patterns of learning analytics implementations emerging in the Australian higher education context challenges the largely essentialist and positivist ontological and epistemological assumptions that underpin many extant learning analytics implementation frameworks (cf. Davenport & Harris, 2007).

Learning Analytics Implementations as Iterative, Dynamic, and Sustainable

Based on these findings, Colvin et al. (2015) generated a model of strategic capability that presents learning analytics as a situated, multidimensional, dynamic, and emergent response to inter-relationships between six mediating dimensions; these not only afford and enable learning analytics implementations, but also recursively shape each other over time (Figure 24.1).

Figure 24.1 presents Colvin et al.'s Model of Strategic Capability learning analytics implementations that represents the phenomena as complex, dynamically interconnected, and temporal, and suggests that actual performance of learning analytics implementations will in turn generate future capacity. In this respect, and observed by the authors, the tenets of the Minimal Viable Product (MVP) (Münch et al., 2013), and the Rapid Innovation Cycle (Kaski, Alamäki, & Moisio, 2014), with their advocacy for an ongoing, iterative, recursive, processual approach to product development and implementation, have traction in the learning analytics implementation space and are recommended to institutional leaders as possible implementation paradigms.

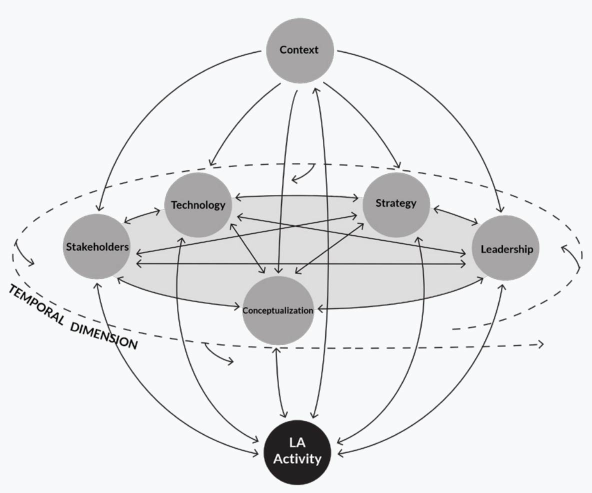


Figure 24.1. Model of Strategic Capability (Colvin et al., 2015, p. 28).

CONCLUSION

Colvin et al.'s (2015) work makes important empirical and methodological contributions to the research literature on learning analytics implementations. First, it provides empirical insights into the relationships between antecedents (affordances) of learning analytics implementations and their outcomes (that is, how they *looked*). Soliciting participants' meanings and understandings of actual learning analytics and learning analytics implementations provided fine-grained, nuanced insight into the varied ways that institutional leaders conceptualized learning analytics implementations, and allowed relationships between these conceptualizations and actual operationalizations to be revealed.

The conceptualization and analysis of learning analytics implementations in Colvin et al.'s (2015) research as multidimensional phenomena resonates with tenets of emerging learning analytics implementation literature (cf. Ferguson et al., 2015; Greller & Drachsler, 2012), and the Model of Strategic Capability generated by their research offers a rich, holistic, systemic conceptualization of learning analytics as a temporal, situated, dynamic consequence of multiple, intersecting, interdependent factors. Of particular significance though is the empirical insight Colvin et al. (2015) afford into the relative salience of the primary sociocultural, technical, and structural factors mediating learning analytics implementations.

Colvin et al.'s (2015) presentation of learning analytics diverges from many extant models, which often frame learning analytics as linear and/or unidimensional phenomena. We suggest that these latter conceptualizations, through their reductionist orientation, do not have the potential to fully capture the complexity, breadth, or disruption of learning analytics implementations, and may be inadvertently militating against the adoption and development of sustainable and effective learning analytics practices and strategies. By contrast, Colvin et al.'s (2015) findings remind us that learning analytics implementations are complex, shaped by interdependent "soft and hard" dimensions (Greller & Drachsler, 2012), and have the potential to challenge and disrupt traditional management and organizational structures in universities. Their research provides institutional learning analytics managers with an empirically derived conceptual framework that highlights the complexity of learning analytics implementations, recognizes the mediating role of context, and can facilitate intra- and inter-institutional evaluation of learning analytics strategies and priorities.

It must be noted that Colvin et al.'s (2015) research has limitations: its data were primarily qualitative, and gleaned from a relatively small sample of institution participants (n=32) located in one higher education context (Australia). Therefore, any direct application of the findings to alternate higher education contexts should be undertaken with some caution. Further, most institutions were found to be in the very early stages of their learning analytics implementations: their programs were embryonic and still developing. Relationships reported between antecedent dimensions and outcomes are therefore to be interpreted within these temporal constraints. It is recommended that further empirical analyses of learning analytics implementations are conducted over time that allow more nuanced insight into this critical area. Notwithstanding, the findings of Colvin et al. (2015) offer an analytics framework that grounds learning analytics implementations in the tenets of multidimensionality and complexity that should have resonance with the broader analytics community.

REFERENCES

Accard, P. (2015). Complex hierarchy: The strategic advantages of a trade-off between hierarchical supervision and self-organizing. *European Management Journal*, 33(2), 89–103.

Arnold, K. E., Lonn, S., & Pistilli, M. D. (2014). An exercise in institutional reflection: The learning analytics readiness instrument (LARI). *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 163–167). New York: ACM.

Arnold, K. E., Lynch, G., Huston, D., Wong, L., Jorn, L., & Olsen, C. W. (2014). Building institutional capacities and competencies for systemic learning analytics initiatives. *Proceedings of the 4th International Conference on Learning Analytics and Knowledge* (LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 257–260). New York: ACM.

Baker, R., & Inventado, P. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics* (pp. 61–75). New York: Springer.

Bichsel, J. (2012). *Analytics in higher education: Benefits, barriers, progress and recommendations*. Louisville, CO: EDUCAUSE Center for Applied Research.

Bolden, R. (2011). Distributed leadership in organizations: A review of theory and research. *International Journal of Management Reviews*, 13(3), 251–269.

Carbonell, K. B., Dailey-Hebert, A., & Gijselaers, W. (2013). Unleashing the creative potential of faculty to create blended learning. *The Internet and Higher Education*, 18, 29–37.

Colvin, C., Rogers, T., Wade, A., Dawson, S., Gašević, D., Buckingham Shum, S., . . . Fisher, J. (2015). *Student retention and learning analytics: A snapshot of Australian practices and a framework for advancement*. Canberra, ACT: Australian Government Office for Learning and Teaching.

Conde, M. Á., & Hernández-García, Á. (2015). Learning analytics for educational decision making. *Computers in Human Behavior*, 47, 1–3.

Daniel, B. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920.

Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Boston, MA: Harvard Business School Press.

Dawson, S., Heathcote, L., & Poole, G. (2010). Harnessing ICT potential. *International Journal of Educational Management*, 24(2), 116–128.

Drachsler, H., & Greller, W. (2012). The pulse of learning analytics understandings and expectations from the stakeholders. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 120–129). New York: ACM.

ECAR-ANALYTICS Working Group. (2015). *The predictive learning analytics revolution: Leveraging learning data for student success ECAR working group paper*. Louisville, CO: EDUCAUSE.

Ferguson, R., Macfadyen, L., Clow, D., Tynan, B., Alexander, S., & Dawson, S. (2015). Setting learning analytics in context: Overcoming the barriers to large-scale adoption. *Journal of Learning Analytics*, 1(3), 120–144.

Foreman, S. (2013a). Five steps to evaluate and select an LMS: Proven practices. *Learning Solutions Magazine*. http://www.learningsolutionsmag.com/articles/1181/five-steps-to-evaluate-and-select-an-lms-proven-practices.

Foreman, S. (2013b). The six proven steps for successful LMS implementation. *Learning Solutions Magazine*. http://www.learningsolutionsmag.com/articles/1214/the-six-proven-steps-for-successful-lms-implementation-part-1-of-2.

Graham, C. R., Woodfield, W., & Harrison, J. B. (2013). A framework for institutional adoption and implementation of blended learning in higher education. *The Internet and Higher Education*, 18, 4–14.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society*, 15(3), 42–57.

Hazy, J. K., & Uhl-Bien, M. (2014). Changing the rules: The implications of complexity science for leadership research and practice. In D. V. Day (Ed.), *Oxford Handbook of Leadership and Organizations*. Oxford, UK: Oxford University Press.

Houchin, K., & MacLean, D. (2005). Complexity theory and strategic change: An empirically informed critique. *British Journal of Management*, 16(2), 149–166.

Kaski, T., Alamäki, A., & Moisio, A. (2014). A multi-discipline rapid innovation method. *Interdisciplinary Studies Journal*, 3(4), 163.

Kotter, J. P., & Schlesinger, L. A. (2008). *Leading Change*. Boston, MA: Harvard Business Review Press.

Laferrière, T., Hamel, C., & Searson, M. (2013). Barriers to successful implementation of technology integration in educational settings: A case study. *Journal of Computer Assisted Learning*, 29(5), 463–473.

Macfadyen, L., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, 15(3), 149–163.

Macfadyen, L., Dawson, S., Pardo, A., & Gašević, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, 9(2), 17–28.

Münch, J., Fagerholm, F., Johnson, P., Pirttilahti, J., Torkkel, J., & Jäarvinen, J. (2013). Creating minimum viable products in industry-academia collaborations. In B. Fitzgerald, K. Conboy, K. Power, R. Valerdi, L. Morgan, & K.-J. Stol (Eds.), *Lean Enterprise Software and Systems* (Vol. 167, pp. 137–151). Springer Berlin Heidelberg.

Norris, D., Baer, L., Leonard, J., Pugliese, L., & Lefrere, P. (2008). Action analytics: Measuring and improving performance that matters in higher education. *EDUCAUSE Review*, 43(1), 42.

Norris, D. M., & Baer, L. L. (2013). *Building organizational capacity for analytics*. Louisville, CO: EDUCAUSE.

Oster, M., Lonn, S., Pistilli, M. D., & Brown, M. G. (2016). The learning analytics readiness instrument. *Proceedings of the 6th International Conference on Learning Analytics and Knowledge* (LAK '16), 25–29 April 2016, Edinburgh, UK (pp. 173–182). New York: ACM.

Owston, R. (2013). Blended learning policy and implementation: Introduction to the special issue. *The Internet and Higher Education*, *18*, 1–3.

Ritchey, T. (2011). General morphological analysis: A general method for non-quantified modelling. In T. Ritchey (Ed.), *Wicked Problems: Social Messes*. http://www.swemorph.com/pdf/gma.pdf

Siemens, G., Dawson, S., & Lynch, G. (2013). *Improving the quality and productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics*. Sydney, Australia: Australian Government Office for Teaching and Learning.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30.

Chapter 25: Learning Analytics: Layers, Loops and Processes in a Virtual Learning Infrastructure

Ruth Crick

Institute for Sustainable Futures, University of Technology Sydney, Australia
Department of Civil Engineering , University of Bristol, United Kingdom
DOI: 10.18608/hla17.025

ABSTRACT

This chapter explores the challenges of learning analytics within a virtual learning infrastructure to support self-directed learning and change for individuals, teams, organizational leaders, and researchers. Drawing on sixteen years of research into dispositional learning analytics, it addresses technical, political, pedagogical, and theoretical issues involved in designing learning for complex systems whose purpose is to improve or transform their outcomes. Using the concepts of 1) layers – people at different levels in the system, 2) loops – rapid feedback for self-directed change at each level, and 3) processes – key elements of learning as a journey of change that need to be attended to, it explores these issues from practical experience, and presents working examples. Habermasian forms of rationality are used to identify challenges at the human/data interface showing how the same data point can be apprehended through emancipatory, hermeneutical, and strategic ways of knowing. Applications of these ideas to education and industry are presented, linking learning journeys with customer journeys as a way of organizing a range of learning analytics that can be collated within a technical learning infrastructure designed to support organizational or community transformation.

Keywords: Complexity, learning journeys, infrastructure, dispositional analytics, self-directed learning

Learning analytics (LA) is a term that refers to the use of digital data for analysis and feedback that generates actionable insights to improve learning. LA feedback can be used in two ways: 1) to improve the personal learning power of individuals and teams in self-regulating the flow of information and data in the process of value creation; and 2) to respond more accurately to the learning needs of others. The growth of new types of datasets that include "trace data" about online behaviour; semantic analysis of human online communications; sensors that monitor "offline" behaviours, locations, bodily functions, and more; as well as traditional survey data collected from individuals, raises significant challenges about the sort of social and technical learning infrastructures that best support processes of improvement, adaptation and change. These challenges are located at the human/data interface and are as important for schools or universities whose purpose is to enhance learning and its outcomes as they are for corporate organizations whose purpose is the provision of a service or a product. Both depend on the capability of humans within their systems to be able to monitor, anticipate, regulate, and adapt productively to complex, rapidly flowing information and to utilize it in their own learning journey to achieve a purpose of value. Learning analytics provides formative feedback at multiple levels in an organization: the same datasets can be aggregated for individuals, teams, and whole organizations. When learning analytics are aligned to shared organizational purposes and embedded in a participatory organizational culture, then new models of change emerge, driven by internal agency and agility, rather than by external regulation.

This chapter reports on a sixteen-year research program of dispositional learning analytics that provided a rich experience in the technical, philosophical, and pedagogical challenges of using data to enhance self-regulated learning at all levels in an organization, rather than simply for "top down" decision making. It utilizes the metaphor of a "learning journey" as a framework for connecting different modes of learning analytics that together constitute a virtual learning ecology designed to serve a particular social vision. Drawing on systems thinking, it uses the themes of "layers," "loops," and "processes" as characteristics of complex learning infrastructures and as a way of approaching the design of learning analytics (Blockley, 2010).

LEARNING ANALYTICS FOR FORMATIVE ASSESSMENT OF LEARNING

The challenge of assessing learning dispositions was the starting point for the research program that began in 2000 at the University of Bristol in the UK. Its rationale was drawn from two findings from earlier research: 1) that data identified and collected for assessment purposes drives pedagogical practice; and 2) that high-stakes summative testing and assessment depresses students' motivation for learning and drives "teaching to the test" (Harlen, 2004; Harlen & Deakin Crick, 2002, 2003). This was a design fault in education systems that have changed little over the last century but that aspire to prepare students for life in the age of "informed bewilderment" (Castells, 2000). The challenge for the research program was first to identify, then to find a way to formatively assess, the sorts of personal qualities that enable people to engage profitably with new learning opportunities in authentic contexts when the "outcome was not known in advance" (Bauman, 2001).

Drawing on a synthesis of two concepts – 1) learning power (Claxton, 1999) and 2) assessment for learning (Broadfoot, Pollard, Osborn, McNess, & Triggs, 1998) – the original factor analytic studies identified seven "learning power" scales and the computation of new variables through which to measure them. These scales included aspects of a person's learning that are both intra-personal as well as inter-personal, drawing on a person's story and cultural context (Wertsch, 1985). Designed as a practical measure, together they were referred to as "dimensions of learning power" (Yeager et al., 2013). The scales were validated with school age students (Deakin Crick, Broadfoot, & Claxton, 2004; Deakin Crick, McCombs, Broadfoot, Tew, & Hadden, 2004; Deakin Crick & Yu, 2008) and with an adult population (Deakin Crick, Haigney, Huang, Coburn, & Goldspink, 2013) and in 2014 the accumulated data (<70K) was re-analyzed to produce a more rigorous and parsimonious instrument, known as the Crick Learning for Resilient Agency Profile (CLARA; Deakin Crick, Huang, Ahmed Shafi, & Goldspink, 2015).

The purpose of the research was to collect data for teachers that would enable them and their students to understand and optimize the processes of learning: specifically to hand over responsibility for change to students themselves by providing them with formative data and a language with which to develop actionable insights into their own learning journeys. The team used technology to generate immediate personalized feedback from the survey data through computing and representing the dimensions of learning power as a spider diagram. This immediate, visual analytic presents the learning power latent variable scores as a pattern to invite personal reflection. Numerical scores are avoided because they represent a form of analytic rationality (Habermas, 1973) more often used to grade and compare for external regulation and encourage entity thinking rather than integral and dynamic thinking (Morin, 2008). The visual analytic provides a framework and a language for a coaching conversation that moves between the individual's identity and purpose and their learning goals and experiences. It provides diagnostic information to turn self-diagnosis into strategies for change (see Figure 25.1). This later became part of the definition of learning analytics (Long & Siemens, 2011; Buckingham Shum, 2012).

Since the first studies were completed, ongoing research explored learning and teaching strategies that enable individuals to become responsible, self-aware learners by responding to their Learning Power profiles (Deakin Crick & McCombs, 2006; Deakin Crick, 2007a, 2007b; Deakin Crick, McCombs, & Haddon, 2007). The focus was on those factors that influence learning power and the sorts of teaching cultures that develop it (Deakin Crick, 2014; Deakin Crick & Goldspink, 2014; Godfrey, Deakin Crick, & Huang, 2014; Willis, 2014; Ren & Deakin Crick, 2013; Goodson & Deakin Crick, 2009; Deakin Crick & Grushka, 2010). Further empirical studies identified pedagogical strategies that support learning power: coaching for learning (Wang, 2013; Ren, 2010), authentic pedagogy (Deakin Crick & Jelfs, 2011; Huang, 2014), teacher development (Aberdeen, 2014), and enquiry-based learning (Deakin Crick, Jelfs, Huang, & Wang, 2011; Deakin Crick et al., 2010). The conceptual framework provided by the dimensions of learning power provides specificity, assessability, and thus practical rigour to an otherwise conceptually and empirically complex social process.

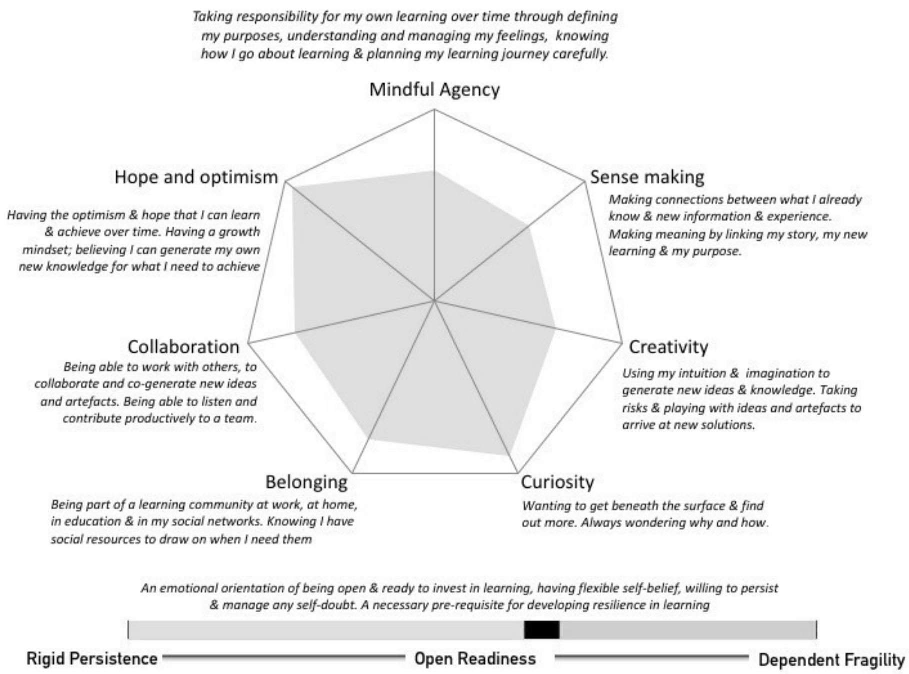


Figure 25.1. Individual Learning Power Profile visual analytic.

The provision of the survey and the calculation of the latent variables in a web-based format afforded the possibility of using the same dataset for rapid feedback for users at four different levels in organizations. First, individual feedback via the visual analytic is available privately to the individual. Second, anonymized mean scores for groups are computed so that a teacher or facilitator can look at the learning characteristics of a teaching group or class and adapt their pedagogy accordingly. Third, anonymized mean scores combined with other variables from an organization's management information system, such as demographics and grouping variables, and others such as attainment or well-being measures, allow for a more sophisticated exploration of data in a whole cohort or organization to inform leadership decision making. Finally, with appropriate permissions, anonymized data can be harvested for exploratory research at a systems level. Each feedback point provides data that carries actionable insights – in other words new learning opportunities (see Figure 25.2).

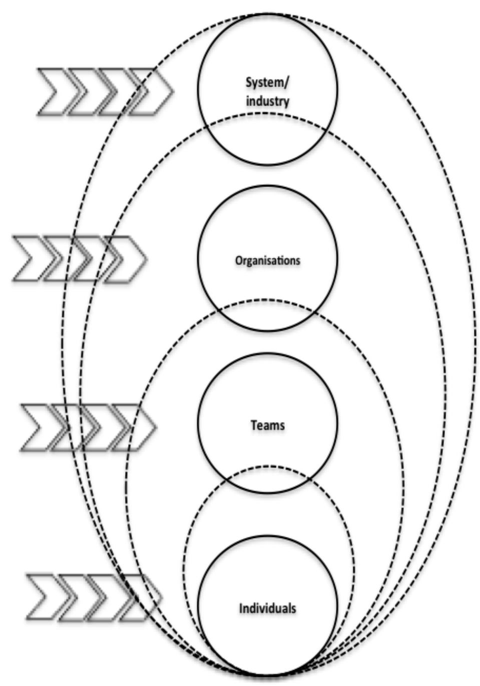


Figure. 25.2. Rapid analytics feedback at four levels.

Rapid feedback of personal data about learning forms part of the emerging field of dispositional learning analytics (Buckingham Shum & Deakin Crick, 2012). It transgresses traditional boundaries, not only in social science, where the focus for research purposes is often on one variable at a time with a view to exploring the impact of one entity on another to serve a research purpose. Traditional boundaries are also transgressed in terms of data systems, which are often understood as the province of system leaders rather than for empowering subjects themselves to become self-directed learners. Traditionally, data is collected for system leaders rather than for individuals. The infrastructure for gathering data at scale, managing stakeholder permissions, and providing a range of analytics from real time summaries to exploratory research has raised challenges in terms of technology, the politics of power, organizational learning agility, and personalization.

LEARNING ANALYTICS AND PEDAGOGY: POINTS OF TENSION

This research program served a practical purpose in the development of pedagogies for personalization and engagement in a variety of educational and corporate settings. The term "dispositional learning analytics" was not used in until as late as 2012 (Buckingham Shum & Deakin Crick, 2012) and the focus on learning analytics was an "unintended outcome" of the original program. The unique interactions between technology, computation, and human learning combined over time to make this approach powerful, sustainable, and scalable in a way that was not possible in assessment practices until the emergence of technology. These interactions are crucial to understanding and developing the emerging field of learning analytics since they also raise significant challenges as well as opportunities.

Perhaps the most significant challenge is in the intrinsic transdisciplinarity of this approach and the need for quality in three different fields – social science, learning analytics, and practical pedagogy, the last of which is highly complex and engages with many forms of human rationalities and relationships. The information explosion has changed forever the ways in which humans relate to information and this adds more complexity (Morin, 2008). These affordances and challenges will be addressed in the next section.

Visual Representation of Data: Making the Complex Meaningful

The presentation of a latent variable in data representing how a person responds to a self-report questionnaire about their learning power is complex. Learning power itself is described as "an embodied and relational process through which we regulate the flow of energy and information over time in the service of a purpose of value" (Deakin Crick et al., 2015, p. 114). For data to be useful to the individual in this context, it needs to be meaningful but sufficiently complex and open to allow for interpretation and response in an authentic context. The goal of learning power assessment is to develop people as resilient agents able to respond and adapt positively to risk, uncertainty, and challenge. As Rutter (2012) argues, the meaning of experience is what matters in resilience studies; resilience is an interactive, "plastic" concept, a state of mind, rather than an intelligence quotient or a temperament. Thus the form in which data about a person's resilient agency is presented to them needs to be fit for this purpose – sufficiently robust to be reliable and valid in traditional terms but sufficiently open-ended and flexible to be "recognized" and responded to in a particular context. The visualization of the spider diagram achieves this goal of being precise enough whilst maintaining a representation of complexity, plasticity, and provisionality. The absence of numbers on the spider diagram is significant since in Western culture a number is often interpreted as an "entity" and not a "process" and can lead to a "fixed" rather than "growth oriented" mindset (Dweck, 2000). Technology makes the visualization of data more effective.

A further observation about the visualization of learning power data is that it connects with different ways of knowing or different "deep seated anthropological interests" (Habermas, 1973, 1989; Outhwaite, 1994). Habermas describes these as "empirical analytical interests," "hermeneutical interests" and "emancipatory interests." Reflecting on "my learning power" begins with a focus on learning identity – Is this like me? Am I this sort of learner? What is my purpose? How do I want to change? These questions connect with emancipatory rationality, or the drive for "autonomy" (Deci & Ryan, 1985) and these forms of rationality are simply not amenable to interpretation through "standardization" since each human being is unique. At best they may be represented through archetypes (Jacobi, 1980).

Image: Metaphor and Story as Carriers of Data

A ubiquitous finding from the studies has been the use of metaphor, story, and image to communicate the meaning of the learning power dimensions and to enable communal dialogue and sense-making. Most notable was the creation of a community story over a year in an Australian Indigenous community. The story was co-constructed with key characters as (sacred) animals that the community had chosen to represent each learning power dimension (latent variable) that they encountered via the learning analytic. The animals locked in Taronga Zoo combined their learning powers to plan a breakout. The story articulated the community's unique cultural history of oppression whilst opening up opportunities for engaging in forms of 21st-century learning as equals in a new paradigm (Goodson & Deakin Crick, 2009; Deakin Crick & Grushka, 2010; Grushka, 2009). Figure 25.3 is an example of one of the graphics represented and used in that context: a representation of a wedge-tailed eagle, which was one of the outcomes of months of community dialogue and debate, before being finally ratified by the local elders. A detailed discussion of this project is beyond the scope of this chapter. The point is that community sense-making, aligned with technology and dispositional analytics, can be represented by images or visualizations, which are locally empowering because they connect with deeper forms of narrative and tradition. Thus they enable educators to engage profitably and meaningfully – and in a time-relevant manner – with the "perezhivanie" – the lived experiences of communities (John-Steiner, 2000; John-Steiner, Panofsky, & Smith, 1994).

Figure. 25.3. A visual graphic representing "strategic awareness."
© Black Butterfly Designs

One Data Point and Different Ways of Knowing

An individual produces a learning power self-assessment survey with rapid feedback designed to stimulate self-directed change. The emancipatory rationality for reflection on self and identity required will often take the form of narrative: it's unique, polyvalent, time bound, and "open-ended" (Brueggemann, 1982). When that individual uses the same data to develop a strategy for moving forwards and achieving a learning purpose, then they are likely to be using "interpretive rationality" or, in Habermasian terms, "hermeneutical rationality." They will be reflecting on a goal and making judgments about the best way to achieve it, collating qualitative data, collaborating with others and drawing on existing funds of knowledge. If they then develop a "measurement model" so that they are able to judge whether they have achieved their purpose, then they will use empirical/strategic rationality – using analytical, means-end logic to determine whether they have succeeded in their purpose.

Thus one useful data point can be apprehended through different rationalities or ways of knowing when it is harnessed to a personal or social purpose. When data is aggregated and anonymized, teachers or facilitators can evaluate and analyze the data quantitatively to assess whether their pedagogy is achieving its purpose, and interpret their findings in order to improve and adjust what they do. Here the same forms of rationality are in operation at an organizational level but the focus shifts to hermeneutical and strategic rationality in the service of a shared purpose. The data is used for leadership decision making. At a systems level, when the same accumulated data is analyzed in a way that is abstracted from context, then the modus operandi is predominantly strategic/analytical rationality.

Rapid Feedback at Multiple Levels: Critical for Ownership and Improvement

Technology enables easy collection, automatic computation, and rapid feedback of survey data to users at different levels in organizations. Where an organization enables self-managing teams to operate in pursuit of shared organizational purpose (Laloux, 2015) then rapid feedback of data that informs both process and outcome is a crucial resource. For example, in a learning organization such as a school or college, a shared purpose is that students develop lifelong learning competences. In this case, CLARA data can be owned and used for improvement by students themselves, by teachers who want to evaluate how effectively their pedagogy is producing lifelong learning competences, by leaders in making decisions about overall college policy in relation to these outcomes, and by researchers who explore and analyze the data to produce new knowledge. Two aspects of this are important: first the rapid feedback and second the sense of ownership of data and professionalism that the feedback affords and requires. A time lag is no longer necessary between data collection, analysis, and feedback of survey data. Historically such time lags often meant that feedback arrived too late to change practice in the context in which data was collected. From a practitioner's perspective, the research was "done to them" rather than owned and responded to by them. Closing this lifecycle gap between research and practice is a critical task to which learning analytics makes a significant contribution.

Top Down or Bottom Up?

Related to this is the sense of participation and ownership in improvement processes that this affords. The authority to interpret the data is aligned with the responsibility to respond to it and improve practice. This has profound implications for societies in which politically defined external regulatory frameworks have become anachronistic and can often work against quality, collaboration, evolution, and transformation. Typically, such frameworks are produced by politicians for accountability purposes and are based on worldview assumptions rooted in the industrial era. Put simply, they often measure the wrong things and the purpose is political accountability and control. So key questions are Who does this data belong to? and Whose purposes are being served? Whilst there is a strong argument for some top-down accountability in learning systems – particularly where young people and public finance are concerned – there is an equally strong argument for empowerment and self-regulation at micro (individual) and meso (organizational) levels. This complexity is a key quality of self-organizing systems that, by definition, require forms of professionalism (commitment to purpose) that go beyond compliance with external regulatory frameworks.

Understanding the system as a whole is a key to learning analytics aimed at improvement (Bryk, Gomez, Grunrow, & LeMahieu, 2015; Bryk, Gomez, & Grunow, 2011; Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010). Borrowing from the world of systems thinking in Health and Industry (Checkland, 1999; Checkland & Scholes, 1999; Snowden & Boone, 2007; Sillitto, 2015) the rigorous analysis of a system leads to the identification of improvement aims and shared purpose – the boundaries of the system are aligned around its purpose (Blockley, 2010; Blockley & Godfrey, 2000; Cui & Blockley, 1990). Thus an alignment around purpose at all levels in a system will both require and enable data to be owned at all levels by those responsible for making the change and using data for actionable insights (see Figure 25.4). The power, or the authorship, of decision-making is both inclusive and participatory. Technical systems and analytics need to reflect this.

Practical Data: Fitness for Purpose

A key issue in learning analytics has to do with the reliability and validity of data, particularly when it is used to judge quality of performance and/or process in authentic contexts. More than ever, quality is an ethical issue. Does this self-assessment tool measure what it purports to measure? Who has the authority to interpret the results? As Yeager et al. (2013) argue, "conducting improvement research requires thinking about the properties of measures that allow an organization to learn in and through practice" (p. 9). They go on to identify three different types of measures that serve three different purposes: 1) for accountability, 2) for theory development, and 3) for improvement. They characterize the latter as practical measures that may measure intermediary targets, framed in a language targeted to specific units of change, contextualized around common experiences and engineered to be embedded in everyday practice. Practical measures may be used for assessing change, for predictive analysis, and for priority setting.

The focus of these practical measures for Yeager and colleagues (2013) is on their use in change programs in organizations led by improvement teams or leadership groups. However, an additional purpose of practical measures is to stimulate ownership, awareness, and responsibility for change on the part of individual users. Practical measures present a theoretical challenge for psychometricians in terms of the need for new summary statistics that contribute to quality assurance where historically much theoretical development has been focused on measures for accountability and theory development, such as internal consistency, reliability, and validity. This issue of authority and responsibility is crucial and one response is to apply a "fitness for purpose" criterion. If the purpose of the measure is to stimulate individual awareness, ownership, and change then the subject of that purpose must have the authority to judge the validity and trustworthiness of the measure since, when it comes to emancipatory rationality and narrative data, the subject is a unique self-organizing system. If the purpose of the measure were to provide government with a measure of success of a policy then the professional community of domain-specific statisticians would have the authority to judge the reliability, validity, and therefore trustworthiness of the data. In both cases, the judgement is about fitness for purpose.

For the emerging field of learning analytics – with its focus on formative feedback for improvement for individuals, teams, and organizations – these issues represent an

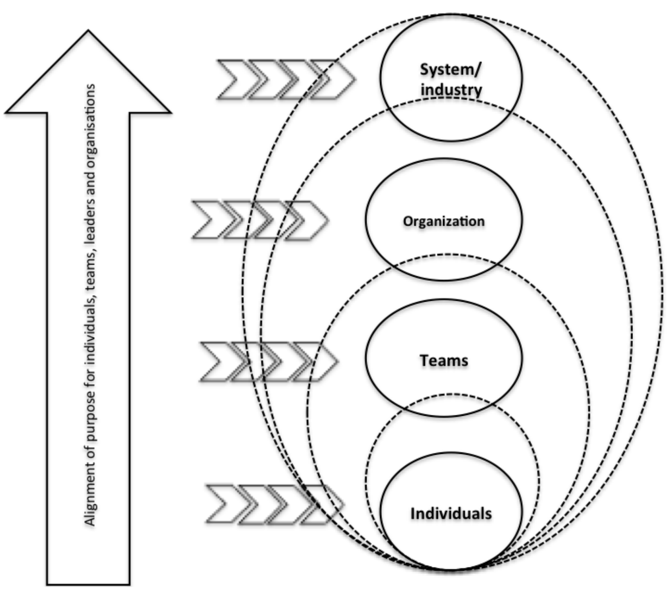


Figure 25.4. Rapid feedback at four levels aligned to organizational purpose.

important field for development. If an assessment tool sold to an institution has no scientific rigour behind it, then even the most sophisticated technology and modes of delivery will not compensate. On the other hand, an organization might want to serve a small number of questions to its community via a tool such as Blue Pulse[1] as a "sensor mechanism" randomized to test the anonymized views of stakeholders about the direction of particular organizational strategies. How do they select the most useful items? What "weight of evidence" do they ascribe to the subsequent data?

In the context of rapid social and technological change, these issues are common. Tools designed for accountability or theory development often have no practical value and thus limited usefulness as learning analytics, whilst tools designed to support practice and improvement often have no theoretical or empirical rigour. The social and moral challenge for learning analytics is to combine and manage both.

TECHNICAL CHALLENGES

The focus of this chapter so far has been on one particular dispositional learning analytic, CLARA, and the issues this has raised. However, any survey tool designed as a practical measure to provide rapid feedback of data for improvement purposes will face similar issues. This section focuses on some of the technical challenges that have framed the experience of the learning power assessment community through successive iterations of technical platforms designed to service the tool.

Survey Platforms

Perhaps because the popular understanding of "surveys" is that they belong to the researchers who administer them, it has been a challenge to build a survey platform to capture data for use at different levels in an organization. The purpose of the data captured in learning analytics is for the subject first, then the facilitator/teacher, then the organization, and finally for researchers, whose brief is to research the whole process or undertake blue skies research on the resultant anonymized datasets. This is because learning can only be done by the learner themselves (Seely Brown & Thomas, 2009, 2010; Thomas & Seely Brown, 2011). This turns the traditional research survey platform on its head. The biggest challenge lies in data protection and ethics and the need for the platform to perform the following functions:

1. Know the identity of each user in order to provide personalized feedback and save that identity for matching with other variables.
2. Protect users at higher "levels" in the organization from the knowledge of the ID of each user whilst matching teachers and students (or employees and managers) where appropriate.
3. Harvest and store anonymized data for research purposes across projects.

What is critical is the underlying data structure that needs to link IDs with three types of variables: demographic, grouping, and survey. Without this flexible data structure, the opportunities presented by learning analytics using self-report survey data are severely limited.

Since 2002, learning power research and development teams have prototyped six platforms. The current solution is through a partnership with one of the world's leading survey providers, whose business development strategy is aligned closely with the vision for learning analytics to support the organizational improvement cycle. The Surveys for Open Learning Analytics (SOLA)[2] platform powered by eXplorance Blue hosts research validated surveys and provides feedback at four levels: for individual users to support personal change; for team leaders to respond more accurately to the learning needs of their groups; for organizational leadership decision making and for systems-wide analysis and research. Examples of feedback at each level are presented in the Appendices.

Collaborative Business Models

This form of collaboration raises issues about the "ownership" of the model amongst the stakeholders who have an interest in it and the management of intellectual property. These include researchers, practitioners, policy makers, and business – both the "education and training" business and the "technology" business. Developing platforms capable of realizing the potential of learning analytics requires business models capable of supporting collaboration, evolution, and innovation and meeting the needs of diverse stakeholders. The interests of different parties have to be balanced in a way that serves the common good, rather than permitting, say, commercial interests to "colonize" research interests, or practitioner interests to "colonize" technical interests. A key factor is the typical lifecycle needs of these differing stakeholders that have to be understood and "harmonized" in order for each to benefit over time.[3]

Identity Management

ID management is a key factor for the learning analytics included in wider virtual ecologies for learning. The challenges of ID management are fundamentally

[1] www.eXplorance.com

[2] www.learningemergence.com

[3] The Learning Emergence network that crowd-sourced the funds for the SOLA platform formed a Limited Liability Partnership in the UK as a vehicle to provide this level of flexibility, linking research, enterprise, and practice around the world: www.learningemergence.com

ethical in nature – protecting individuals' personal data, providing feedback that is personal and supported in appropriate ways through coaching and learning relationships where needed whilst enabling stakeholders at different levels in the system to access anonymized data where they have permission and to match datasets where that is required by research, so as to explore the patterns and relationships between variables in complex learning infrastructures.

Apart from surveys, many tools are learning analytics in the sense that they provide formative feedback for individuals to support their learning in some way. For example, the Assessment of Writing Analytics being developed by (Selvin & Buckingham Shum, 2014) or the iDapt[4] tool for reflexive understanding of an individual's mental models that shape their approach to their professional task (Goldspink & Foster, 2013). The former is a way of critiquing and supporting individuals in developing academic argumentation as a part of their knowledge generation whilst the latter addresses issues of identity and purpose through repertory grid analysis (Kelly, 1963). These are both critical aspects of learning journeys whose focus is on feedback for awareness, ownership, and responsibility for the process of learning on the part of the individual.

LEARNING ANALYTICS AND LEARNING JOURNEYS

Using learning analytics to stimulate change in learning power inevitably invites questions about the wider ecology of processes and relationships that can empower individuals to adapt profitably to new learning opportunities. This is particularly important in authentic contexts where the outcome is rarely known in advance. The metaphor of a "learning journey" was adopted to reflect the complex dynamics of a learning process that begins with forming a purpose and moves iteratively towards an outcome or a performance of some sort. Learning power enables the individual or team to convert the energy of purpose into the power to navigate the journey, to identify and select the information, knowledge, and data they need to work with to achieve that purpose (see Figure 25.5; Deakin Crick, 2012). When an individual or a team learns something without reflecting on the process of learning at a meta-level, this is "single loop" learning. Double loop learning is when the individual or team is able "reflexively to step back" from the process and learn how to learn with a view to improving the process and doing it more effectively next time. They are intentionally becoming more agile and responsive in regulating the flow of information and data that they need to achieve their purpose.

[4] www.inceptlabs.com.au

This framework provides a useful model through which to understand the sort of learning infrastructure and the analytics to support learning and improvement. It provides a typology for learning analytics tuned to support learning and, critically, to enable learners to step back from the "job they are doing" to reflect at a meta-level, to monitor, anticipate, respond, and learn – in other words to engage in double loop learning (Bateson, 1972; Bruno, 2015) as depicted in Figure 25.6.

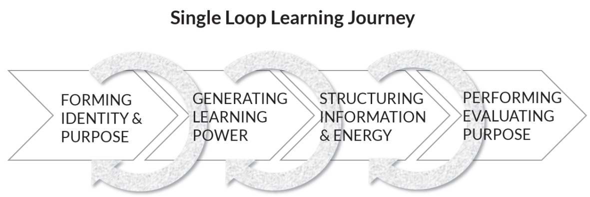


Figure. 25.5. A single loop learning journey.

The Four Processes of a Learning Journey

A learning journey is a dynamic whole with distinguishable sub-processes. It has a natural lifecycle and is collaborative as well as individual, personal as well as public. Learning journeys happen all of the time at different levels and stages. Learning is framed by a purpose, an intention, or a desire that provides the "lens" through which the individual or team can identify and focus on the information that matters. Articulating purpose is the first stage of the "meta language" of learning. Without purpose, learning lacks direction and discipline and it is difficult to select from a welter of data the information that really matters. Developing personal learning power through which to articulate a purpose and respond to data is the second process. The third is the structuring and re-structuring of information necessary to achieve the particular purpose. The final process is the production and evaluation of the product or performance that achieves the original the purpose.

A learning journey is an intentional process through which individuals and teams regulate the flow of energy and information over time in order to achieve a purpose of value. It is an embodied and relational process, which can be aligned and integrated at all levels in an organization, linking purpose with performance and connecting the individual with the collective. The learning power of individuals and teams converts the potential energy of shared purpose into change and facilitates the process of identifying, selecting, collecting, curating, and constructing knowledge in order to create value and achieve a shared outcome.

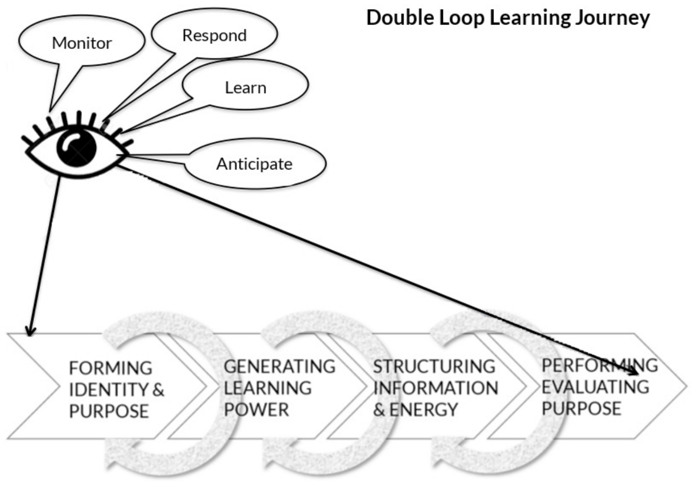


Figure 25.6. Double loop learning journeys.

Learning Journeys as Learning Infrastructure

It is no longer sensible or even possible to separate the embodied and the virtual in learning. The future is both intensely personal and intensely technological. The challenge is to align social and personal learning journeys with the sorts of technologies and learning analytics that serve them and scaffold intentional "double loop" learning.

A learning journey has a generic architecture: it has stages that occur in a sequence, a start event, a finish event, and many transition events in between. It covers a particular domain; it faces inwards as well as outwards; it is framed by user need or purpose. It is iterative and cumulative. It is focused on a stakeholder or customer task and is ideally aligned to organizational target outcomes. Each stage has many "next best actions" and interactions, framed by a meta-movement between purpose and performance. Stakeholders use personal learning power and knowledge structuring tools to navigate their journey. Stages have transition rules and interaction rules and stakeholders can be on many journeys at the same time. A journey can be collaborative or individual, simple or complex, high value or low value. Journeys follow stakeholders across whatever channels they choose and they are adaptive to individual behaviour. They cover different territories with domain-specific sets of knowledge and know-how and they integrate knowing, doing, and being.

There are at least three distinct applications of these ideas for learning analytics:

1. To frame the relational, social, and technical learning infrastructure of an organization so that individuals and teams become more agile, responsive, and able to respond productively to change and innovation
2. To design models that explore and explain stakeholder behaviour – how students or customers embody purpose, learning power, knowledge, and performance in order to communicate more accurately with stakeholder communities
3. To develop digital infrastructure to support self-directed learning and behaviour change, at scale, in particular domains – in other words, mass education, across domains as defined and embracing as, for example, climate change or financial competence

Learning Infrastructure for Living Labs: Learning Analytics for Learning Journeys

To be resourced at scale, learning journeys require a network infrastructure that accesses information and experience from a wide range of formal and informal sources, inside and beyond the organization. The individual or team relates to all of these in identifying, selecting, and curating what they need in terms of 1) information and data and 2) "how to go about it" expertise for achieving their purpose. This network infrastructure is part of a wider ecosystem with platforms that scaffold these relationships drawing on cloud technology, mobile technology, social learning and curating, learning analytics for rapid feedback, "big data" and badges – to name some key analytic genres.

There are strong synergies here with best practices in real-time, omni-channel customer communication management through the recognition that many customer communication journeys are in fact learning journeys. They also involve individuals in engaging with information in order to achieve a purpose (Crick & MacDonald, 2013).

Providing and servicing this learning infrastructure is a specialist task that is arguably the purpose of a "living lab" or a "network hub" in an improvement community. The network (social and organizational relationships) and the ecosystem (technical resources to scaffold these relationships) provide an infrastructure with permeable boundaries between research, policy, practice, and commercial enterprise, facilitating engaged, trans-disciplinary, carefully structured improvement prototypes. Such a "hub" is sometimes described as a "living lab," the purpose of which is to integrate engaged, user-driven improvement research with technology, professional learning, and the wider research community. It provides, and researches, core hub functions and expertise in partnership with the enterprises needed to scale and deliver learning services.

Supporting such a learning infrastructure requires sustained attention to different types of expertise and resource development, including the following:

- The personal and social relationships necessary for facilitating and leading learning journeys including storying, reflection, personal learning power, and purposing
- The organizational arrangements that support learning journeys as a modus operandi for improvement – such as rapid prototyping, coaching, and agile learning cultures
- The architecture of space (virtual and embodied) within the relevant domain of service
- The technologies, tools, and analytics that support the processes of learning journeys through rapid feedback of personal and organizational data for stimulating change, defining purpose, knowledge structuring, and value management
- The virtual learning ecosystem that facilitates and enhances participatory learning relationships across the project/s at all levels – users, practitioners, and researchers

Figure 25.7 presents a high-level design for such a learning journey infrastructure.[5]

A Transition in Thinking

The idea of a learning journey is simple and intuitive. The metaphor facilitates an understanding of learning as a dynamic process; however, it does represent a fundamental transition in how we understand knowledge, learning, identity, and value. Knowledge is no longer a "stock" that we protect and deliver through relatively fixed canons and genres; it is now a "flow" in which we participate and generate new knowledge, drawing on intuition and experience. Its genres are fluid and institutional warrants are less valuable (Seeley Brown, 2015). Learning power is the way in which we regulate that flow of energy and information over time in the service of a purpose of value – rather than a way of receiving and remembering "fixed" knowledge from experts. Millennial identity is found not in ownership and control, but in creating, sharing, and "remixing" – in agency, impact, and engagement. Value is generated in the movement between purpose and performance. Leadership is about learning our way forward together.

What Next?

A plethora of candidate tools and platforms use learning analytics to optimize and support learning for individuals and to improve learning contexts. Tools that address reflective writing (Simsek et al., 2015), sense making , coaching, knowledge curation and sharing (www.declara.com), harnessing collective intelligence (Buckingham Shum, 2008; Buckingham Shum & De Liddo, 2010; Buckingham Shum & Ferguson, 2010), and leadership decision making (Barr 2014) to name a few.

The big challenge for 21st-century learning professionals is understanding how these tools and platforms cohere within a learning organization, a virtual collaboratory or a living lab, in which the focus is on the learning of a whole community of interest, such as those concerned with renewing a city's infrastructure, or a geographical region, or wide domains of public concern such as financial education or climate change. Many of the ideas and learning analytics practices discussed here have been developed and applied in different contexts already. What is required next is a way of integrating these ideas and practices in an authentic and grounded context, focusing on how the whole fits and flows together. This requires a business model for all stakeholders that makes collaboration – not competition – the modus operandi. It requires all stakeholders to abandon "silos" in favour of networks, and be willing to "learn a way forward together," which inevitably also means having permission to fail. In short, these ideas form a starting point for the resourcing of a living lab or network hub, supported by a partnership of core providers constructing coherence from their four contextual perspectives: research, industry, the business world of "learning professionals," and the personal learning of all stakeholders.

CONCLUSION

This chapter has focused on some of the challenges and opportunities of the use of technology and computation for enhancing the *processes* of learning and improvement – rather than only the outcomes. Learning analytics and the affordances of technology have become game-changers for sustainability for organizations in a data-rich, rapidly changing world. Learning analytics are designed to provide formative feedback at multiple levels and these can be aggregated

[5] This high level learning journey infrastructure is derived from customer journey architecture developed in the financial services market by Decisioning Blueprints Ltd. www.decblue.com

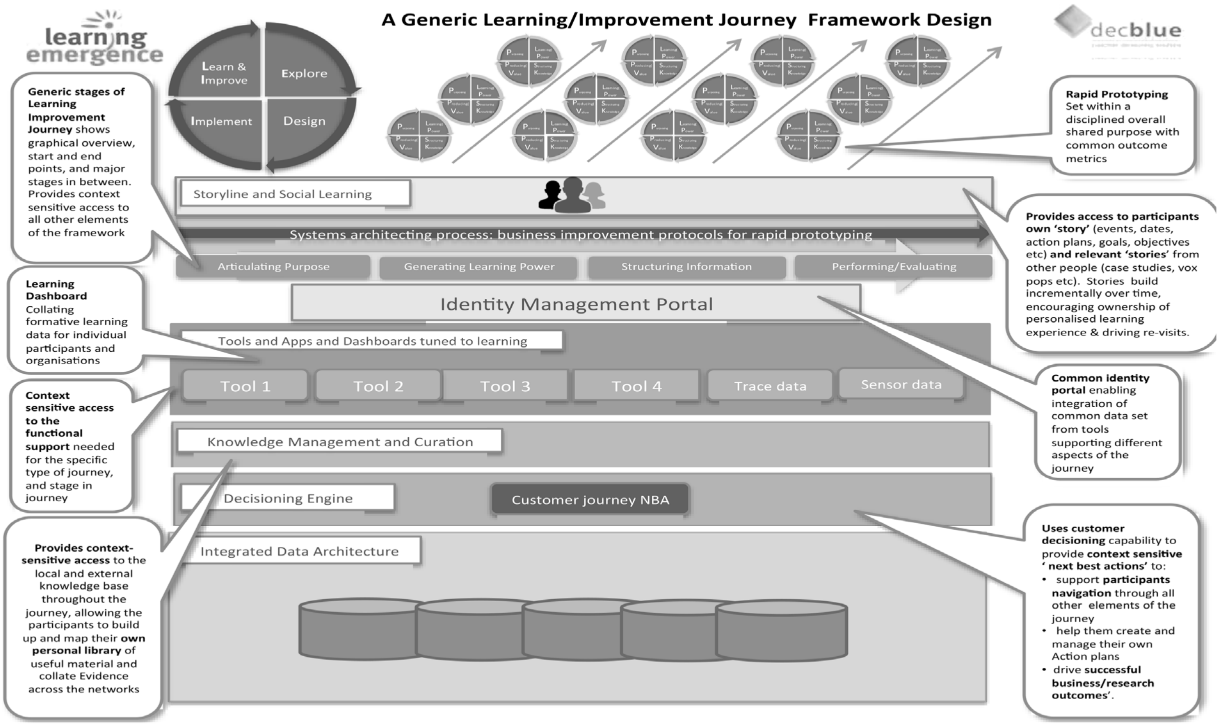


Figure 25.7. High level design for learning architecture.
© Decisioning Blueprints Ltd

for individuals, teams, and whole organizations. When learning analytics are aligned to shared organizational purposes and embedded in a participatory organizational culture, new models of change emerge capable of integrating external regulation with internal agency and agility. This chapter began with an account of a learning analytic that focused on generating learning power for individuals. As its research and development program progressed, it became clear that learning power and its associated analytics were just one part of a more complex and dynamic learning journey. The learning journey is a useful metaphor for framing the way we think about and design learning analytics, as part of the sort of learning infrastructure we need to develop, for learning organizations and in wider social contexts such as living labs. The technical, political, commercial, and philosophical challenges are immense and can only be met by thinking and design that account for complexity and participation.

REFERENCES

Aberdeen, H. (2014). *Learning power and teacher development.* Graduate School of Education Bristol, UK.

Barr, S. (2014). An integrated approach to decision support with complex problems of management. University of Bristol, UK.

Bateson, G. (1972). *Steps to an ecology of mind.* San Francisco, CA: Chandler.

Bauman, Z. (2001). *The individualized society.* Cambridge, UK: Polity Press.

Blockley, D. (2010). The importance of being process. *Civil Engineering and Environmental Systems*, 27(3), 189–199.

Blockley, D., & Godfrey, P. (2000). *Doing it differently: Systems for rethinking construction.* London: Telford.

Broadfoot, P., Pollard, A., Osborn, M., McNess, E., & Triggs, P. (1998). Categories, standards and instrumentalism: Theorising the changing discourse of assessment policy in English primary education. https://eric.ed.gov/?id=ED421491

REFERENCES

Brueggemann, W. (1982). *The creative word: Canon as a model for biblical education.* Philadelphia, PA: Fortress Press.

Bruno, M. (2015). A *foresight review of resilience engineering: Designing for the expected and unexpected.* London: The Lloyds Register Foundation.

Bryk, A., Gomez, L., Grunrow, A., & LeMahieu, P. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard, MA: Harvard Educational Press.

Bryk, A., Gomez, L. M., & Grunow, A. (2011). Getting ideas into action: Building networked improvement communities in education. In M. T. Hallinan (Ed.), *Frontiers in sociology of education* (pp. 127–162). Springer Netherlands.

Bryk, A., Sebring, P., Allensworth, E., Luppescu, S., & Easton, J. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago, IL: University of Chicago Press.

Buckingham Shum, S. (2008). Cohere: Towards web 2.0 argumentation. Paper presented at the 2nd International Conference on Computational Models of Argument, 28–30 May 2008, Toulouse, France.

Buckingham Shum, S. (2012). *Learning analytics.* Moscow: UNESCO Institute for Information Technologies.

Buckingham Shum, S., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (LAK '12), 29 April–2 May 2012, Vancouver, BC, Canada (pp. 92–101). New York: ACM.

Buckingham Shum, S., & De Liddo, A. (2010). Collective intelligence for OER sustainability. Paper presented at the 7th Annual Open Education Conference, 2–4 November 2010, Barcelona, Spain.

Buckingham Shum, S., & Ferguson, R. (2010). Towards a social learning space for open educational resources. Paper presented at the 7th Annual Open Education Conference, 2–4 November 2010, Barcelona, Spain.

Castells, M. (2000). *The rise of the network society*, vol. 1 of The information age: Economy, society and culture. Oxford, UK: Blackwell.

Checkland, P. (1999). *Systems thinking, systems practice.* John Wiley.

Checkland, P., & Scholes, J. (1999). *Soft systems methodology in action.* Chichester, UK: Wiley.

Claxton, G. (1999). *Wise up: The challenge of lifelong learning.* London: Bloomsbury.

Crick, T., & MacDonald, E. (2013). *One customer at a time.* Cranfield University.

Cui, W., & Blockley, D. I. (1990). Interval probability theory for evidential support. *International Journal of Intelligent Systems*, 5(2), 183–192.

Deakin Crick, R. (2007a). Enquiry based curricula and active citizenship: A democratic, archaeological pedagogy. *CitizEd International Conference Active Citizenship*, Sydney.

Deakin Crick, R. (2007b). Learning how to learn: The dynamic assessment of learning power. *Curriculum Journal*, 18(2), 135–153.

Deakin Crick, R. (2014). Learning to learn: A complex systems perspective. In R. Deakin Crick, C. Stringer, & K. Ren (Eds.), *Learning to learn: International perspectives from theory and practice*. London: Routledge.

Deakin Crick, R., Broadfoot, P., & Claxton, G. (2004). Developing an effective lifelong learning inventory: The ELLI project. *Assessment in Education*, 11(3), 248–272.

Deakin Crick, R., & Goldspink, C. (2014). Learner dispositions, self-theories and student engagement. *British Journal of Educational Studies*, 62(1), 19–35.

Deakin Crick, R., & Grushka, K. (2010). Signs, symbols and metaphor: Linking self with text in inquiry based learning. *Curriculum Journal*, 21(1), 447–464.

Deakin Crick, R., Haigney, D., Huang, S., Coburn, T., & Goldspink, C. (2013). Learning power in the workplace: The effective lifelong learning inventory and its reliability and validity and implications for learning and development. *International Journal of Human Resource Management*, 24, 2255–2272.

Deakin Crick, R., Huang, S., Ahmed Shafi, A., & Goldspink, C. (2015). Developing resilient agency in learning: The internal structure of learning power. *British Journal of Educational Studies*, 63(2), 121–160.

Deakin Crick, R., & Jelfs, H. (2011). Spirituality, learning and personalisation: Exploring the relationship between spiritual development and learning to learn in a faith-based secondary school. *International Journal of Children's Spirituality*, 16(3), 197–217.

Deakin Crick, R., Jelfs, H., Huang, S., & Wang, Q. (2011). *Learning futures final report*. London: Paul Hamlyn Foundation.

Deakin Crick, R., Jelfs, H., Symonds, J., Ren, K., Grushka, K., Huang, S., Wang, Q., & Varnavskaja, N. (2010). *Learning futures evaluation report*. University of Bristol, UK.

Deakin Crick, R., & McCombs, B. (2006). The assessment of learner centred principles: An English case study. *Eudcational Research and Evaluation*, 12(5), 423–444.

Deakin Crick, R., McCombs, B., Broadfoot, P., Tew, M., & Hadden, A. (2004). *The ecology of learning: The ELLI two project report*. Bristol, UK: Lifelong Learning Foundation.

Deakin Crick, R., McCombs, B., & Haddon, A. (2007). The ecology of learning: Factors contributing to learner centred classroom cultures. *Research Papers in Education*, 22(3), 267–307.

Deakin Crick, R., & Yu, G. (2008). Assessing learning dispositions: Is the effective lifelong learning inventory valid and reliable as a measurement tool? *Educational Research*, 50(4), 387–402.

Deci, E., & Ryan, R. (1985). *Intrinsic motivation and self-determination in human behaviour*. New York: Plenum.

Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. New York: Psychology Press.

Godfrey, P., Deakin Crick, R., & Huang, S. (2014). Systems thinking, systems design and learning power in engineering education. *International Journal of Engineering Education*, 30(1), 112–127.

Goldspink, C., & Foster, M. (2013). A conceptual model and set of instruments for measuring student engagement in learning. *Cambridge Journal of Education*, 43(3), 291–311.

Goodson, I., & Deakin Crick, R. (2009). Curriculum as narration: Tales from the children of the colonised. *Curriculum Journal*, 20(3), 225–236.

Grushka, K. (2009). Identity and meaning: A visual performative pedagogy for socio-cultural learning. *Curriculum Journal*, 20(3), 237–251.

Habermas, J. (1973). *Knowledge and human interests*. Cambridge, UK: Cambridge University Press.

Habermas, J. (1989). *The theory of communicative action: The critique of functionalist reason*, vol. 2. Cambridge, UK: Polity Press.

Harlen, W. (2004). A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes. *Research Evidence in Education Library*, EPPI-Centre, Social Science Research Unit, Institute of Education, London.

Harlen, W., & Deakin Crick, R. (2002). A *systematic review of the impact of summative assessment and testing on pupils' motivation for learning*. Evidence for Policy and Practice Co-ordinating Centre, Department for Education and Skills, London.

Harlen, W., & Deakin Crick, R. (2003). Testing and motivation for learning. *Assessment in Education*, 10(2), 169–207.

Huang, S. (2014). Imposed or emergent? A critical exploration of authentic pedagogy from a complexity perspective. University of Bristol, UK.

Jacobi, J. (1980). *The psychology of C. G. Jung*. London: Routledge & Kegan Paul.

John-Steiner, V. (2000). *Creative collaboration*. Oxford, UK: Oxford University Press.

John-Steiner, V., Panofsky, C., & Smith, L. (1994). *Socio-cultural approaches to language and literacy: An interactionist perspective*. New York: Cambridge University Press.

Kelly, G. A. (1963). *A theory of personality*. New York: Norton.

Laloux, F. (2015). *Reinventing organisations: A guide to creating organizations inspired by the next stage of human consciousness*. Nelson Parker.

Long, P., & Siemens, G. (2011). Penetrating the fog: Learning analytics and education. *Educause Review*, September/October, 31–40. https://net.educause.edu/ir/library/pdf/ERM1151.pdf

Morin, E. (2008). *On complexity*. New York: Hampton Press.

Outhwaite, W. (1994). *Habermas: A critical introduction*. Cambridge, UK: Polity Press.

Ren, K. (2010). "Could do better, so why not?" Empowering underachieving adolescents. University of Bristol, UK.

Ren, K., & Deakin Crick, R. (2013). Empowering underachieving adolescents: An emancipatory learning perspective on underachievement. *Pedagogies: An International Journal, 8*(3), 235–254.

Rutter, M. (2012). Resilience as a dynamic concept. *Development and Psychopathology*, 24, 335–344.

Seely Brown, J. (2015). Re-imagining libraries and learning for the 21 century. Aspen Institute Roundtable on Library Innovation, 10 August 2015, Aspen, Colorado. http://www.johnseelybrown.com/reimaginelibraries.pdf

Seely Brown, J., & Thomas, D. (2009). Learning for a world of constant change. Paper presented at the 7th Gillon Colloquium, University of Southern California.

Seely Brown, J., & Thomas, D. (2010). Learning in/for a world of constant flux: Homo sapiens, homo faber & homo ludens revisited. In L. Weber & J. J. Duderstadt (Eds.), *University Research for Innovation: Glion VII Colloquium* (pp. 321–336). Paris: Economica.

Selvin, A., & Buckingham Shum, S. (2014). *Constructing knowledge art: An experiential perspective on crafting participatory representations*. San Rafael CA: Morgan & Claypool.

Sillitto, H. (2015). *Architecting systems: Concepts, principles and practice*. London: College Publications Systems Series.

Simsek, D., Sandor, A., Buckingham Shum, S., Ferguson, R., De Liddo, A., & Whitelock, D. (2015). Correlations between automated rhetorical analysis and tutors' grades on student essays. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March 2015, Poughkeepsie, NY, USA (pp. 355–359). New York: ACM. doi:10.1145/2723576.2723603

Snowden, D., & Boone, M. E. (2007, November). A leader's framework for decision making. *Harvard Business Review*. https://hbr.org/2007/11/a-leaders-framework-for-decision-making

Thomas, D., & Seely Brown, J. (2011). *A new culture of learning: Cultivating the imagination for a world of constant change*. CreateSpace Independent Publishing Platform.

Wang, Q. (2013). Coaching psychology for learning: A case study of enquiry based learning and learning power development in secondary education in the UK. University of Bristol, UK.

Wertsch, J. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.

Willis, J. (2014). Learning to learn with indigenous Australians. In R. Deakin Crick, C. Stringher, & K. Ren (Eds.), *Learning to learn: International perspectives from theory and practice* (pp. 306–327). London: Routledge.

Yeager, D., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013, December). *Practical measurement*. San Francisco, CA: Carnegie Foundation for the Advancement of Teaching. https://www.carnegiefoundation.org/resources/publications/practical-measurement/

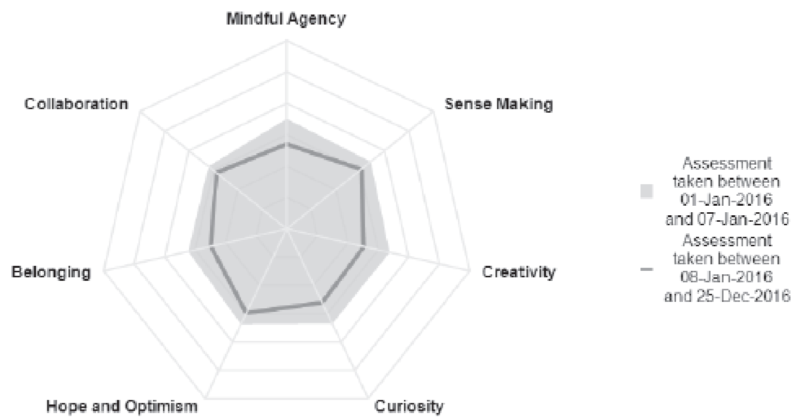
APPENDIX I

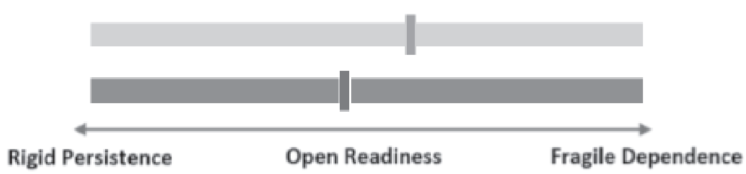
learning emergence

CLARA Group Overview

Creation Date: **19 May 2016**
Group Name: **Class Room - 4 A**

Here is the collective Learning Power Profile of Class Room - 4 A



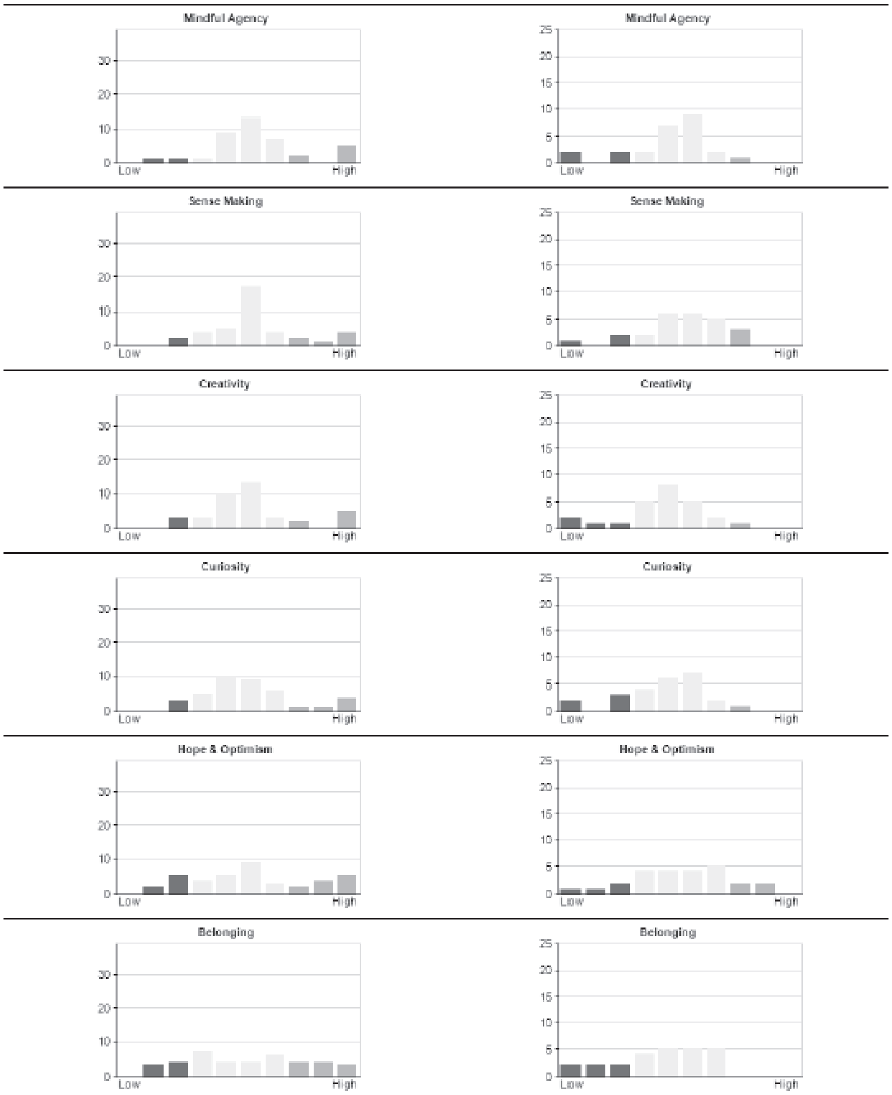


	Assessment taken between 01-Jan-2016 and 07-Jan-2016	Assessment taken between 08-Jan-2016 and 25-Dec-2016
Sub-group size	40	40
Responses Received	39	25
Response rate	97%	62%

CLARA Group Overview for Class Room - 4 A 1/3
Report Generated on 19 May 2016

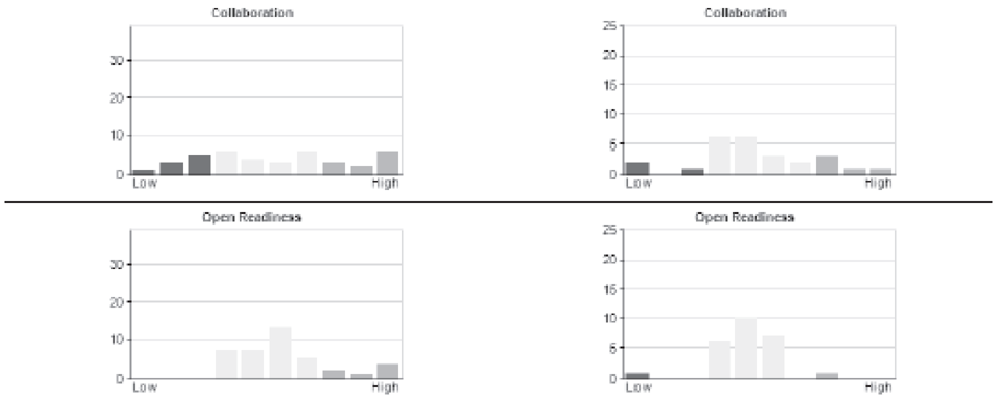
The histograms below compare the score distribution in the first and second assessment in each of the eight CLARA dimensions:

Assessment taken between 01-Jan-2016 and 07-Jan-2016	Assessment taken between 08-Jan-2016 and 25-Dec-2016



CLARA Group Overview for Class Room - 4 A
Report Generated on 19 May 2016

2/3



This table below compares numerically the score distributions in the first and the second assessment oh this group:

CLARA Dimension	Assessment	N	Min.	Max.	Mean	S.D.
Belonging	1	39	12.28	100.00	53.77	24.30
	2	25	0.00	69.18	41.32	19.53
Collaboration	1	39	4.42	100.00	53.26	27.88
	2	25	0.00	95.58	47.84	23.01
Creativity	1	39	26.22	100.00	56.17	19.78
	2	25	0.00	74.46	42.29	18.21
Curiosity	1	39	26.90	100.00	55.85	19.60
	2	25	0.00	74.48	43.44	18.14
Hope and Optimism	1	39	13.72	100.00	56.17	24.73
	2	25	0.00	86.28	49.30	21.17
Mindful Agency	1	39	16.50	99.88	58.44	19.16
	2	25	0.00	77.62	44.90	18.29
Orientation to Learning	1	39	30.36	100.00	56.93	18.93
	2	25	0.00	79.44	44.84	13.46
Sense Making	1	39	27.32	100.00	57.21	19.13
	2	25	0.00	76.06	50.85	17.38

1=Assessment taken between 01-Jan-2016 and 07-Jan-2016
2=Assessment taken between 08-Jan-2016 and 25-Dec-2016

N=Response Received
Min.= the lowest score reported
Max.= the highest score reported
S.D. = standard deviation of the score distribution

Chapter 26: Unleashing the Transformative Power of Learning Analytics

Linda L. Baer[1], Donald M. Norris[2]

[1]Senior Fellow, Civitas Learning, USA
[2]President, Strategic Initiatives, Inc., USA
DOI: 10.18608/hla17.026

ABSTRACT

Learning analytics holds the potential to transform the way we learn, work, and live our lives. To achieve its potential, learning analytics must be clearly defined, embedded in institutional processes and practices, and incorporated into institutional student success strategies. This paper explains both the interconnected concepts of learning analytics and the organizational practices that it supports. We describe how learning analytics practices are currently being embedded in institutional systems and practices. The embedding processes include changing the dimensions of organizational culture, changing the organizational capacity and context, crafting strategies for student success, and executing change management plans for student success. These embedding processes can dramatically accelerate the improvement of student success at institutions. The paper includes some institutional success stories and instructive failures. For institutions seeking to unleash the transformative power of learning analytics, the key is an aggressive combination of leadership, active strategy, and change management.

Keywords: Learning transformation, 21st century skills, change management, competence building, employment skills, predictive learning analytics, personalized learning

Some keen observers of higher education profess that learning analytics-based practices hold the potential to transform traditional learning (Ali, Rajan, & Ratliff, 2016). They can also transform competence building that focuses on skills for employment (Weiss, 2014). The combination of predictive learning analytics, personalized learning, learning management systems, and effective linkages to career and workforce knowledge can dramatically shape the manner in which we prepare for and live our lives. Learning analytics will be a linchpin in this multifaceted transformation.

Across the higher education landscape, learning analytics practices are growing. Individual faculty, learning analytics experiments, innovations, and pilot projects – the academic innovation equivalent of "1,000 points of light" – are demonstrating the value of learning analytics in practice (Sclater, Peasgood, & Mullan, 2016, p. 15). Faculty are building experience in deploying and improving learning analytics practices and sharing their knowledge with their colleagues. Groups like the Bill & Melinda Gates Foundation have actively sponsored so-called "next gen" learning projects to advance such efforts.[1] When they prove their success, these innovations are being nurtured to evolve into full-blown institutional initiatives. Collaborative efforts like the Open Academic Analytics Initiative (OAAI) has developed and deployed an open-source academic early-alert system that can predict (with 70–80% accuracy) within the first 2–3 weeks of a term which students in a course are unlikely to complete the course successfully (Little et al., 2015). Eventually, such innovations may spread across the higher education and knowledge industry.

Learning analytics practices are also shaping a new generation of academic technology infrastructure. Literally every enterprise resource planning (ERP) and learning management systems (LMSs) vendor is embedding analytics in its products and services

1 See Next Generation Learning Challenge: http://www.educause.edu/focus-areas-and-initiatives/teaching-and-learning/next-generation-learning-challenges

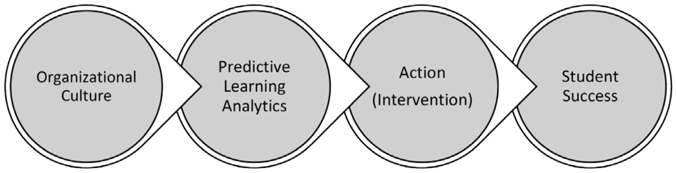


Figure 26.1. Position of predictive learning analytics in the path of student success.

(Baer & Norris, 2015). Consortia like UNIZIN (Hilton, 2014) have emerged to develop an in-the-cloud, next generation digital learning environment (NGDLE) that can accommodate loosely coupled learning applications, learning object repositories, personalized learning, and analytics capabilities. Learning analytics is a key issue in next gen technology infrastructures and processes.

Clearly, the elements of ubiquitous, predictive learning analytics are coalescing. Practitioners are striving to understand their meaning and how to leverage their impact. Individual faculty and staff, institutional leaders, policy makers, and employers all face different opportunities and challenges in confronting the emergence of learning analytics. In our work with institutions, we have confronted the following questions from various stakeholder groups and individual practitioners:

- How can individual faculty become accomplished learning analytics practitioners, building their expertise and elevating learning analytics practice, across courses and majors, among their peers and across their institution?
- How can institutional faculty and staff involved with student success initiatives embed learning analytics to support dynamic interventions and actions?
- How can institutional leaders and policy makers develop supportive policies, practices, and learning/course management tools that unleash the transformative power of learning analytics across their institutions and beyond?
- How can employers and policy makers provide effective linkages to career and workforce knowledge that will further enhance and extend student success to include academic, co-curricular development, employability, and career elements?

This chapter will help illuminate these opportunities and challenges and provide the context for enabling each of these different stakeholders to understand how to capitalize on the potential of learning analytics.

EMBEDDING LEARNING ANALYTICS IN INSTITUTIONAL SYSTEMS, PRACTICES, AND STUDENT SUCCESS STRATEGIES

In *The Predictive Learning Analytics Revolution*, the ECAR-Analytics Working Group developed a highly simplified model of the student success process (Little et al., 2015) illustrated in Figure 26.1. It features the central position of predictive learning analytics and action/intervention in the middle of the student learning process. Analytics is essential to informed action/intervention. Without action, analytics is merely reporting; and without an analytics-based foundation, interventions are actions shaped imperfectly by instinct and belief. Enhancing student success depends on this progression of predictive learning analytics to action, all in the context of organizational culture.

The ECAR-Analytics Working Group goes on to stipulate that "before deploying predictive learning analytics solutions, an institution should ensure that its organizational culture understands and values data-informed decision making processes. Equally important is that the organization be prepared with the policies, procedures and skills needed to use the predictive learning analytics tools and be able to distill actionable intelligence from their use" (Little et al., 2015, p. 3).

Changing the Dimensions of Organizational Culture

This is a laudable prescription. However, our experience with many institutions suggests that even when student success initiatives are thoughtfully launched, organizational culture does not change rapidly, or all at once. In reality, student success initiatives are characterized by the continuous, parallel evolution of organizational culture, organizational capacity, and specific student success projects and actions. Such campaigns often require five to seven years of implementation before yielding substantial organizational change results (Norris & Baer, 2013). Moreover, understanding the *dimensions* of the change needed can shift over time; as well, institutional teams, through experience and reflection, develop a better understanding of analytics-enabled student success interventions in practice.

Table 26.1. Changing Organizational Culture for Student Success in the 21st Century

Dimension Of Culture	Traditional Institutional Culture	Transformed Institutional Culture
Use of Data in Decision Making	Culture of Reporting	Culture of Evidence/Performance Imperative of Knowing
Innovation	1,000 Points of Light	Successful Innovations are Taken to Scale
Collaboration	Individual Faculty Autonomy	Student Success is Everyone's Job Community of Practice
Scope of Student Success	Academic Achievement and Development	Fully Integrated Success Assessment: Academic/Curricular Achievement, Co-Curricular Development, Work-Related Experiences, DIY Competence Building, and Other Elements

As President Miyares of University of Maryland University College points out, "Often absent from the dialogue is an acknowledgment of the heavy lifting required to leverage analytics as a strategic enabler to transform an institution. There is no 'easy button' for improving the financial, educational, and operational outcomes across an institutional enterprise. Doing so requires a combined commitment of technology, talent, and time to help high-performing colleges and universities leverage analytics not only for one-time insights but also for ongoing performance management and improvement guided by evidence-based decision making" (Miyares & Catalano, 2016).

Table 26.1 illustrates this point. The first dimension of cultural change needed to enable predictive learning analytics-based interventions for student success involves the use of data in decision making. Institutions must change from a culture of reporting – with no imperative to action – to a culture of performance-based evidence and a commitment to action. This dimension is obvious to most new student success teams. But it takes time and experience with new approaches for the change to take hold and become an embedded practice.

The second and third dimensions of culture change – innovation and collaboration – are also critical. Most institutions of higher education support innovations with a lower case "I," leaving them up to individual faculty and celebrating 1,000 points of light. But for student success initiatives to be optimized across the institution they must practice Innovation with a "capital I." They must learn how to take successful innovations and interventions to scale, building on success, and achieving consistency in responses and interventions (Norris & Keehn, 2003). The traditional collaboration culture is based on individual faculty autonomy, which reinforces the individual innovation culture. Optimizing student success requires greater collaboration, not only among and between faculty, but involving academic support and administrative staff and cross-disciplinary perspectives. Institutions with the most successful student success initiatives (ECAR, 2015) make student success everyone's job, and dramatically increase the network of supporting and intervening persons across the institution. This can include forming active communities of practice dealing with student success.

The fourth dimension of cultural change involves the scope of student success. While traditionally student success focuses on academic achievement, a 21st century transformed perspective on student success scope expands to include an integrated perspective of academic/curricular achievement, co-curricular development, work-related experiences, and do-it-yourself (DIY) competence building. The transformation of this fourth dimension is developing more slowly than the first three, but it will accelerate when new mechanisms emerge to share workforce competence knowledge and integrate records of learner's demonstrated learning and competences.

Changing Organizational Context/Capacity

It's not just about culture; it's about all aspects of organizational capacity for analytics and student success initiatives. Table 26.2 portrays the five dimensions of organizational capacity, assessed for a sample institution. Institutional teams that undertake enhancing or optimizing student success and making it an institutional priority have found it useful to assess the current state of capacity development (also called a readiness/maturity index), then establish the targets needed to enhance student success over a reasonable planning period, say five years. Table 26.2 illustrates the current capacity score, the targeted score in five years, and the "gap" between the two that needs to be closed over time (the black bars in the "Target Score" column). Institutional leaders need to focus on setting stretch goals for student success capacity that will position them to meet their targets.

The Community College Research Center (CCRC) studied five Integrated Planning and Advising Services (IPAS)

Table 26.2. Institutional Context/Capacity for Analytics and Student Success, Current and Targeted

Dimension of Organizational Context/Capacity	Current Score 1 2 3 4 5	Current Score 1 2 3 4 5
Leadership		
1. Top management is committed to enhancing student success and views predictive learning analytics as essential to student success	4	5
2. Analytics has a senior-level champion who can remove barriers, champion funding	3	5
3. Top leadership is committed to and consistently practices evidence-based decision making	2	4
4. Appropriate funding and investment has been made in analytics, iPASS,* and student success	3	5
Culture/Behaviour		
1. Our institution's culture favours performance-based evidence for decision making	2	3
2. Our culture recognizes "the imperative of knowing" and we practice "action analytics"	2	4
3. We assess student learning and success innovations and take successful innovations to scale	1.5	5
4. We emphasize collaboration in student success efforts and make student success everyone's job	1	2
5. We define and integrate student success to include curricular, co-curricular, work, and talent development	1	2
Technology Infrastructure/Tools/Applications		
1. Capacity to: Store/access disparate data in raw/transformed form	3	4
Store/access predictive results	2.5	4
Deploy/measure the effects of learning interventions	2.5	4
Integrate numerous predictive analytics tools	2	3
2. Computing power for regular big data analyses, simulations, visualizations, and processes	2	3
3. Security protocols in place to ensure learning analytics effort is not a liability	2	3
4. Data governance yields adequate data quality	2.5	3.5
5. Integrate and unify data sources	1	3.5
6. Adequate iPASS Infrastructure to support analytics-driven interventions	3	4.5
Policies, Processes, and Practices		
1. Institutional policies and data stewardship fulfill federal, state, and local laws	1.5	2.5
2. Workflows for student success processes are well documented	2	4
3. Guiding coalition and cross-disciplinary teams for student success	1.5	4
4. Fully integrated planning, resourcing, execution, and communication (PREC) for student success (eliminates fragmentation — "connects the dots")	2	4
Skills and Talent Development		
1. Student success innovation/collaboration skills	1.5	4
2. Specific LA Skills: Data Science: Data analysis, interpretation and visualization	1.5	4
Programming/Vendor product for data mining	2.5	4
Data Literacy — necessary for predictive models/algorithms	2.5	4
Research expertise and understanding of nuanced data	2	4
Intervention — time, frequency, tone, and nature	2	4
Instructional Design — for embedded predictive analytics	1.5	4
3. Capacity for reinvention of student life cycle processes (end-to-end) is well developed	2	3

Legend: 1. Initial; 2. Emerging; 3. Functional; 4. Highly Functional; 5. Exemplary

Source: Adapted from the Norris/Baer framework, ECAR Maturity Index, ECAR-Analytics Working Group, & Educause Maturity and Deployment Indices (Dahlstrom, 2016).

* Individual planning and advising for student success (iPASS) systems combine student planning tools, institutional tools, and student services. iPASS "gives students and administrators the data and information they need to plot a course toward a credential or degree, along with the ongoing assessment and nudges necessary to stay on course toward graduation. iPASS combines advising, degree planning, alerts, and interventions to help students navigate the path toward a credential. These tools draw on predictive analytics to help counselors and advisors determine in advance whether a student is at risk of dropping out or failing out and it can help assist in selecting courses" (Yanosky & Brooks, 2013). iPASS, used in conjunction with the LMS, is emerging as a key mechanism for delivering the analytics-informed interventions that are proving critical to student success.

participants to determine readiness for technology adoption (RTA). The RTA framework "is particularly focused on ensuring that technology-based reforms lead to end-user adoption and changed practice" (Karp & Fletcher, 2014, p. 13). For this to occur, colleges must not only have sufficient technological resources, they must also attend to the cultural components of readiness. Notably, the framework acknowledges the existence of various micro-cultures within an organization — groups of individuals, each with their own culture, norms, and attitudes toward technology (see Karp & Fletcher, 2014).

Crafting Strategies for Student Success

Strategy is focused, consistent behaviour over time, responding to ongoing changes in the environment (Mintzberg, Ahlstrand, & Lampel, 1998). In order to unleash the transformative power of learning analytics, institutions should craft and execute active strategies that enhance student success. These strategies become the mechanism for enhancing capacity and focusing attention on the strategic intent of enhancing student success. Active strategies achieve four outcomes: 1) set direction, 2) focus effort, 3) define the organization, and 4) provide consistency (Baer & Norris, 2016a). Table 26.3 illustrates typical student success strategies for a sample institution.

Change Management Plan for Student Success

Optimizing student success is one of the great change management challenges facing institutional leaders. Institutions need to embark on ongoing expeditions to leverage predictive learning analytics (and others) for student success. Table 26.4 illustrates the sort of overarching change management plan that enables institutions to turn student success initiatives into institutional strategies that will transform culture, processes, and practices over time.

Effective change management interventions for student success can accelerate the rate at which institutions embed learning analytics-driven interventions into the organization. They can focus the attention of leadership at all levels on the importance of changing culture and communication.

INSTRUCTIVE FAILURES AND SUCCESS STORIES

What can be learned from examining institutions that have been embedding learning analytics into their processes, practices, and student success strategies? Let's use two lenses: "instructive failures" and "success stories."

Instructive Failures

What constitutes an instructive failure in the embedding and leveraging of analytics in institutional processes and culture? In most cases, failure does not mean complete and abject rejection of embedded learning analytics. Rather, it means failure to overcome orga-

Table 26.3. Crafting Active Strategies for Student Success

Strategies	Description
Strategy #1: Develop Unified Data, Information, and Predictive Learning Analytics for Student Success	Strategy #1 should focus on improving the data, information, and analytics capacity of the institution. The individual goals under this strategy should focus on technology infrastructure; policies, processes, and practices; and skills and talent development as indicated in Table 26.2. These goals should establish metrics and stretch targets for the five-year planning timeframe.
Strategy #2: Integrate Planning, Resourcing, Execution, and Communication (PREC) for Student Success	Practical experience has shown that institutions are plagued by fragmented processes and practices for planning, resourcing, executing, and communicating (PREC) student success initiatives. Strategy #2 should establish the goal of integrating PREC activities across the seven dimensions of student success optimization: 1) managing the pipeline; 2) eliminating bottlenecks, copying best practices; 3) enabling dynamic intervention, 4) enhancing iPASS; 5) leveraging next gen learning and learning analytics; 6) achieving unified data; and 7) extending the definition of student success to include employability and career success (Baer & Norris, 2016b). This strategy should establish metrics for improving practice along these seven dimensions and set stretch goals.
Strategy #3: Advance Individual Planning and Advising for Student Success (iPASS)	iPASS is one of the game changers in student success. Strategy #3 should focus on enhancing the institution's current iPASS platform and practices. Goals should focus on integrating iPASS with other platforms and analytics and with new approaches to learning interventions such as personalized learning. Dramatically increasing the number, effectiveness, and targeting of interventions should be a goal and metric. See Seven Things You Should Know About IPAS and iPASS Grant Recipients.
Strategy #4: Integrate Personalized Learning and Competence Building into Institutional Practice	Personalized learning also promises to be a game changer. Strategy #4 should focus on expanding next gen learning innovations and taking them to scale across the institution. Introducing competence building will be an important differentiator for the institution in the longer term.
Strategy #5: Integrate Employability and Workforce Data into Institutional Practice	Strategy #5 has a longer time frame than the others but it is likely to have important long-term impacts. Getting started now will enable institutions to establish a competitive advantage.

Table 26.4. Change Management Plan for Student Success

Element	Sample Description
I. Aligning the student success initiatives to institutional context	
Collaboration: What sort of guiding coalition and campus teams are needed?	The institution forms a guiding coalition to oversee student success. This is a cross-disciplinary team that evolved from integrated planning for student success. The guiding coalition will serve as the steward/shepherd for the execution of the student success strategy.
Culture: How can you align with institutional culture and intentionally change it over time?	In assessing its current organizational culture and context, the institution assesses its institutional culture for using data, information, and analytics to support student success. They also assess the current capacity and culture to engage all staff and faculty in student success efforts. They express their intent to move from a culture of reporting to a culture of evidence-based decision making and more aggressive interventions to build student success.
Leadership: What role does leadership at all levels play in optimizing student success?	The institution's assessments establish that executive leadership is critical to building support for student success efforts, mobilizing energies, and building commitment. This is true at all stages of student success development. To achieve a highly functional state of student success achievement, leadership and talent must be developed at all levels of the organization.
II. Connecting new student success initiatives to current student success efforts and data systems/analytics	
Integration: How can we "connect the dots" linking all student success initiatives?	To overcome the extreme fragmentation in its student success processes, the institution participates in integrated strategic planning for student success. Utilizing the rubrics, strategies, and expedition maps emerging from this process, the guiding coalition will assure the integration and alignment of resourcing, execution, and communication.
Data and Analytics Resources: What current resources are available and what additional ones are needed?	The initial gap analysis of investment in student success initiatives and data, information, and analytics, in particular, reveal a performance gap and actions to fill the gap. Leveraging analytics is critical to optimizing student success and to monitoring and setting stretch goals.
Responsibility: Who is responsible for the elements of optimizing student success?	The initial assessment yields a mapping of responsibilities for all aspects of student success optimization. This mapping then guides the formation, composition, and functioning of the guiding coalition.
III. Engaging faculty, staff, and other stakeholders to change their perspectives and practices, and enable process and practice improvement	
Communication: Who are the key stakeholders and what kind of communication plan is needed?	The initial integrated planning assessment yields a mapping of the stakeholders for student success activities. This mapping has guided the formation, composition, and functioning of the guiding coalition.
Talent Development: What sort of talent development is needed to develop faculty and staff?	Talent development needs emerge from innovation and expeditionary strategy crafting workshops. These drive the skills development elements of the five strategies.
Demonstrate Success: How do we achieve early victories and demonstrate value-added and ROI?	So-called "low-hanging fruit" are identified throughout the innovation and expeditionary strategy crafting workshops. These elements are then regularly updated and utilized by the guiding coalition.

nizational inertia to deploy and leverage embedded analytics in an effective manner that aggressively advances student success. Consider the following common examples:

- **Many institutions make poor selection decisions in analytics tools, applications, and solutions.** This may be due to who actually was responsible for purchasing the technology, better solutions that subsequently became available, how the campus systematically determined the integration of the tool, and ongoing investment in people and resources to launch and sustain the technologies. Even some of the most successful institutions have had to overcome poor selection decisions and/or migrate to better options that became available.
- **Fundamentally, many institutions have not invested sufficiently in their data and information foundation.** Their data are fragmented and cannot be combined across siloed databases; data are literally "hiding in plain sight." Achieving unified, accessible data requires persistent attention to data governance and is critical to leveraging analytics to achieve student success.
- **Even institutions that have invested in excellent analytics packages often sub-optimize their use of these capabilities.** By failing to prepare staff and faculty for how they can use these tools to identify at-risk behaviour and launch interventions, institutions sub-optimize their impact. Studies report that faculty believe they could be better instructors and students indicate that they could improve learning if they increased the use of the LMS (Dahlstrom, Brooks, & Bichsel, 2014). All of the ERP LMS and analytics vendors report on this problem, but do not yet provide adequate talent development at the implementation stage to overcome it.
- **Many institutions that do achieve advances in data, information, and analytics often fail to "connect the dots" by integrating all of the differ-**

ent student success interventions and activities across the institution. Failing to integrate planning, resourcing, execution, and communication (PREC) for student success will doom institutional efforts to sub-optimization.

Examples of similar instructive failure experiences can be found at most colleges and universities that are moving forward, but tentatively, in the effective deployment of embedded learning analytics. What can we learn from institutions that excel in leveraging embedded learning analytics?

Success Stories: Institutions Getting it Right

A small group of institutions have displayed the vision, leadership, and careful execution needed to leverage embedded analytics to advance student success. But even these leaders are far from done. They all report that student success analytics efforts are 5–7 year campaigns where the standards for success are continuously shifting upward. These long slogs require ongoing campus learning, flexibility, and research on what works and what doesn't. Consider the following success stories.

American Public University System (APUS) is an online, for-profit provider that has been a leader in using embedded analytics for over ten years. Analytics-driven interventions are part of their culture, and their student success innovations span the entire institution. Every week they evaluate and rank the risk level for all of their students and drive appropriate actions/interventions (Rees, 2016).

Arizona State University (ASU) has been a leader in the use of analytics-driven interventions in remedial education, advising, degree planning, and other vectors of student success over a long period. Their president is a nationally recognized champion in the strategic use of analytics. They have been pursuing adaptive courseware pilots, run as part of the Next Generation Courseware Challenge funded by the Bill & Melinda Gates Foundation, which provide strong evidence of its positive effect on the learner experience. Important to the success was the development of data and dashboards that enhance interaction and communication between faculty and students (Johnson & Thompson, 2016).

University of Maryland University College (UMUC) has developed an industry-leading capability to use predictive analytics to lead targeted learner interventions. Their enterprise-wide effort enjoys strong presidential leadership. UMUC is an industry leader in analyzing big data to identify appropriate and effective learner interventions, and the Center for Innovation in Learning and Student Success (CILSS) is providing research support for these efforts to improve student outcomes. The university is bringing learning analytics to bear on the full range of student concerns, from selecting courses to making the best use of study time.[2]

Sinclair Community College has been investing in data, information, and analytics for well over a decade. Its efforts began with merging three units to create a Business Intelligence Competency Center, developing a comprehensive data warehouse to provide "a single, unified version of the truth," and practice active data stewardship, data integration, and data quality. These supported one of the first iPASS systems and aggressive degree planning, which were used to make analytics-driven intervention in support of student success a key activity across the institution (Moore, 2009).[3]

Colorado State University (CSU) has achieved significant increases in graduation rates and has all but eliminated the minority achievement gap over the past decade. Its strategy for student success calls for even greater gains over the next decade. CSU has enhanced its performance by making student success a recognized institutional strategy and has organized around that principle. They have a VP for Student Success and have mobilized an active network of faculty, staff, and others to improve academic, co-curricular, and other student experiences. Student success is accepted as being everyone's job. Data, information, and analytics have been key elements of the university's progress and are embedded in the highly integrated student success processes and practices (Lamborn & Thayer, 2014).

Even the most advanced institutions do not believe they are done. New tools and practices keep raising the bar for student success analytics and practices. The best is yet to come.

CONCLUSION

As institutions seek to unleash the transformative power of learning analytics, they must be prepared for an extended campaign, punctuated by carefully planned victories and demonstrations of how leveraging analytics can enhance student success. Institutional strategy for student success is an emergent pattern of focused, consistent behaviour over time, responding to changing environmental conditions and new trends as they present themselves (Mintzberg et al., 1998, p. 5). The key to optimizing student success is an aggressive combination of leadership, active strategy, and change management, as illustrated in

[2] See Predictive Analytics Leads to Targeted Learner Interventions: http://www.umuc.edu/innovatelearning/initiatives/analytics.cfm and Predictive Analytics for Student Success: Developing Data-Driven Predictive Models of Student Success: http://www.umuc.edu/visitors/about/ipra/upload/developing-data-driven-predictive-models-of-student-success-final.pdf

[3] See My Academic Plan: https://www.sinclair.edu/services/basics/academic-advising/my-academic-plan-map/

Figure 26.2.

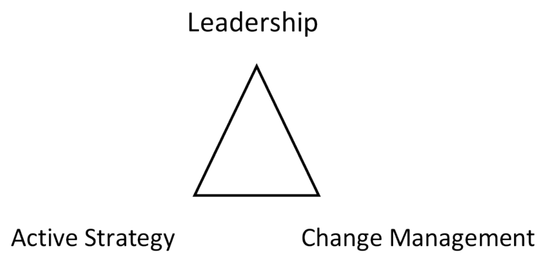


Figure 26.2. Unleashing the transformative power of learning analytics.

To prepare for and leverage these developments, the faculty and administration should coordinate and focus their activities in distinctive ways, using research-based decision making and developing their own site-specific best practices.

Individual Faculty Seeking to Become Accomplished Learning Analytics Practitioners

Learning-analytics-based projects are not just technical innovations; they are adaptive innovations, requiring active engagement of all participants in co-creating the design and outcomes (Heifetz, 2014). Individual faculty should realize that the introduction of learning analytics tools challenge some of the basic, traditional cultures of institutions. They will require new, more collaborative approaches to innovation and to the pervasive use of shared, performance-based evidence. Building skills in learning analytics should be a highly strategic move for faculty – if such skills are recognized, valued, and rewarded by academic leadership.

Communities of learning analytics practice will likely emerge and draw individual faculty into collaborations beyond their academic departments. This will also require a more comprehensive understanding and use of learning and course management systems. ECAR's report on *The Current Ecosystem of Learning Management Systems in Higher Education: Student, Faculty, and IT Perspectives* (Dahlstrom et al., 2014) concluded from surveys that:

- Faculty and students value the LMS as an enhancement to teaching and learning experiences, but relatively few use the full capacity of the systems.
- User satisfaction is highest for basic LMS features and lowest for collaboration and engagement features.
- Faculty say they could be more effective instructors and students could be better students with more skilled use of the LMS.
- Students and faculty want the LMS to have enhanced features and operational functions; be personalized; and use analytics to enhance learning outcomes.

In addition, faculty must understand the emerging features of next generation digital learning environments (Brown, Dehoney, & Millichap, 2015). These new, cloud-based platforms will be loosely coupled and will enable better integration of analytics, varying modes of learning, learning objects, and mobile learn apps.

Institutional Faculty and Staff Involved with Student Success Initiatives

Large-scale student success initiatives are creating new opportunities for individual faculty and staff to participate in cross-department and cross-function collaborations. These efforts will dramatically increase the number and effectiveness of interventions to enhance student success. Greater understanding of what interventions are working for which students will be critical. Institutions can expand the return on investments in interventions when they target the initiatives to students who will most benefit in a timely manner (Baer & Norris, 2016b). There are several approaches to improving student success initiatives including a more focused effort through a student success team. This approach brings multiple institutional players together to better integrate and collaborate on behalf of services to students. Service providers can better leverage institutional resources, jointly reviewing and selecting technologies to support services and providing ongoing evaluation of what works. [4]

More specific to learning analytics is understanding the following: 1) metrics to measure learning, 2) availability, use, and training on tools such as learning management systems and course management systems, 3) an inventory of interventions or actions available as risky learning behaviour is identified, 4) clear policies and practices supported by data governance, and 5) ongoing linkages to skills, competencies, and workforce.

Institutional Leaders and Policy Makers Seeking to Unleash the Transformative Power of Learning Analytics

Institutional leadership must actively discharge their responsibility to craft the institutional strategies and change management plans to enhance student success that will ultimately unleash the power of learning analytics. In the process, they will progressively transform the organizational culture and context. These efforts will require a dynamic combination of leadership,

[4] See Elevation through Collaboration: Successful Interventions for Students on Probation: http://www.nacada.ksu.edu/Resources/Academic-Advising-Today/View-Articles/Elevation-through-Collaboration-Successful-Interventions-for-Students-on-Probation.aspx Hints on how to scale initiatives can be found in Scaling Community College Interventions: http://www.publicagenda.org/files/CuttingEdge2.pdf

active strategy, and change management to achieve their potential. The details are critical in policy and practice development around learning analytics. It is apparent in conversations with campuses that interpretations about data privacy and protection can in fact hinder the adoption and deployment of learning analytics. Institutions must deal with the ecosystem of ethical issues and policy changes in order to 1) insure proper collection and use of data, 2) protect individual rights concerning the use of data, and 3) support the ethical and legal use of data within the institution. To this end, the purpose and boundaries regarding the use of learning analytics should be well defined and visible. Students should be engaged as active agents in the implementation of learning analytics (e.g., informed consent, personalized learning paths, and interventions.) Slade identifies gaps around policy for use of learning analytics in the following areas: 1) moral purpose; 2) purpose and boundaries; 3) informed consent; 4) collection, analyses, access to, and storage of data; 5) students as active agents, and 6) labelling and stereotyping (Slade, n.d.).

Employers and Policy Makers Provide Effective Linkages to Career and Workforce Knowledge

The addition of competence, career, and workforce knowledge to the information base supporting student success efforts will be a major breakthrough. It will stimulate new approaches and practices and support the emergence of highly effective iPASS in institutions. These linkages are continuing to be developed. Some fields are clearly mandating certification and use evidence of skill competencies. Advances in the field are now enabling match ups of job or career needs with competencies demonstrated.

Put simply, learning analytics are destined to be instrumental in the transformation of how we prepare for and live our lives, guided and sustained by perpetual learning. They will surely impact the transformation of many aspects of colleges, universities, and other learning enterprises – and faculty, staff, students, and others who make them work.

REFERENCES

Ali, N., Rajan, R., & Ratliff, G. (2016, March 7). How personalized learning unlocks student success. Educause Review. http://er.educause.edu/articles/2016/3/how-personalized-learning-unlocks-student-success

Baer, L., & Norris, D. M. (2015, October 23). What every leader needs to know about student success analytics: Civitas White Paper. https://www.civitaslearningspace.com/what-every-leader-needs-to-know-about-student-success-analytics/

Baer, L., & Norris, D. M. (2016a). A call to action for student success analytics: Optimizing student success should be institutional strategy #1. https://www.civitaslearningspace.com/StudentSuccess/2016/05/18/call-action-student-success-analytics/

Baer, L., & Norris, D. M. (2016b). SCUP annual conference workshop: Optimizing student success should be institutional strategy #1. Society for College and University Planning (SCUP-51), 9–13 July 2016, Vancouver, BC, Canada.

Brown, M., Dehoney, J., & Millichap, N. (2015, June 22). What's next for the LMS? *Educause Review*. http://er.educause.edu/ero/article/whats-next-lms.

Craig, R., & Williams, A. (2015, August 17). Data, technology, and the great unbundling of higher education. *Educause Review*. http://er.educause.edu/articles/2015/8/data-technology-and-the-great-unbundling-of-higher-education

Dahlstrom, E. (2016, August 22). Moving the red queen forward: Maturing analytics capabilities in higher education. *Educause Review*. http://er.educause.edu/articles/2016/8/moving-the-red-queen-forward-maturing-analytics-capabilities-in-higher-education

Dahlstrom, E., Brooks, D. C., & Bichsel, J. (2014). *The Current Ecosystem of Learning Management Systems in Higher Education: Student, Faculty, IT Perspectives*. ECAR. https://net.educause.edu/ir/library/pdf/ers1414.pdf

ECAR (2015, October 7). The predictive learning analytics revolution: Leveraging learning data for student success. ECAR Working Group Paper. https://library.educause.edu/~/media/files/library/2015/10/ewg1510-pdf.pdf

Heifetz, R. A. (2014, February 18). Adaptive and technical change. http://www.davidlose.net/2014/02/adaptive-and-technical-change/

Hilton, J. (2014, September 15). Enter UNIZIN. *Educause Review*. http://er.educause.edu/articles/2014/9/enter-unizin

Johnson, D., & Thompson, J. (2016, February 2). Adaptive Courseware at ASU. https://events.educause.edu/eli/annual-meeting/2016

Karp, M. M., & Fletcher, J. (2014, May). Adopting new technologies for student success: A readiness framework. http://ccrc.tc.columbia.edu/publications/adopting-new-technologies-for-student-success.html

Lamborn, A., & Thayer, P. (2014, June 17). Colorado State University Student Success Initiatives: What's Been Achieved; What's Next. Enrollment and Access Division Retreat.

Little, R., Whitmer, J., Baron, J., Rocchio, R., Arnold, K., Bayer, I., Brooks, C., & Shehata, S. (2015, October 7). *The predictive learning analytics revolution: Leveraging learning data for student success*. ECAR Working Group Paper. https://library.educause.edu/resources/2015/10/the-predictive-learning-analytics-revolution-leveraging-learning-data-for-student-success

Mintzberg, H., Ahlstrand, B., & Lampel, J. (1998). *Strategy Safari: A guided tour through the wilds of strategic management*. New York: The Free Press.

Miyares, J., & Catalano, D. (2016, August 22). Institutional analytics is hard work: A five-year journey. *Educause Review*. http://er.educause.edu/articles/2016/8/institutional-analytics-is-hard-work-a-five-year-journey

Moore, K. (2009, November). Best practices in building analytics capacity. Action Analytics Symposium.

Norris, D., & Baer, L. (2013). *A toolkit for building organizational capacity for analytics*. Strategic Initiatives, Inc.

Norris, D., & Keehn, A. K. (2003, October 31). IT Planning: Cultivating innovation and value. *Campus Technology*. https://campustechnology.com/Articles/2003/10/IT-Planning-Cultivating-Innovation-and-Value.aspx

Rees, J. (2016, August 30). Adoption leads to outcomes at American Public University System. *Civitas Learning Space*. https://www.civitaslearningspace.com/adoption-leads-outcomes-american-public-university-system/

Sclater, N., Peasgood, A., & Mullan, J. (2016, April 19). Learning analytics in higher education: A review of UK and international practice. https://www.jisc.ac.uk/reports/learning-analytics-in-higher-education

Slade, S. (n.d.). Learning analytics: Ethical issues and policy changes. The Open University (PowerPoint presentation). https://www.heacademy.ac.uk/system/files/resources/5_11_2_nano.pdf

Weiss, M. R. (2014, November 10). Got skills? Why online competency-based education is the disruptive innovation for higher education. *Educause Review*. http://er.educause.edu/articles/2014/11/got-skills-why-online-competencybased-education-is-the-disruptive-innovation-for-higher-education

Yanosky, R., & Brooks, D. C. (2013, August 30). Integrated planning and advising services (IPAS) research. *Educause*. https://library.educause.edu/resources/2013/8/integrated-planning-and-advising-services-ipas-research

Chapter 27: A Critical Perspective on Learning Analytics and Educational Data Mining

Rita Kop[1], Helene Fournier[2], Guillaume Durand[2]

[1]Faculty of Education, Yorkville University, United Kingdom
[2]Information and Communications Technologies, National Research Council of Canada, Canada
DOI: 10.18608/hla17.027

ABSTRACT

In our last paper on educational data mining (EDM) and learning analytics (LA; Fournier, Kop & Durand, 2014), we concluded that publications about the usefulness of quantitative and qualitative analysis tools were not yet available and that further research would be helpful to clarify if they might help learners on their self-directed learning journey. Some of these publications have now materialized; however, replicating some of the research described met with disappointing results. In this chapter, we take a critical stance on the validity of EDM and LA for measuring and claiming results in educational and learning settings. We will also report on how EDM might be used to show the fallacies of empirical models of learning. Other dimensions that will be explored are the human factors in learning and their relation to EDM and LA, and the ethics of using "Big Data" in research in open learning environments.

Keywords: Educational data mining (EDM), big data, algorithms, serendipity

The past ten years have been interesting in the fields of education and learning technology, which seem to be in flux. Whereas past research in education related to the educational triangle of learner, instructor, and course content (Kansanen & Meri, 1999; Meyer & Land, 2006), newly developed technologies put an emphasis on other dimensions influencing learning; for instance, the learning context or learning setting and the technologies being used (Bouchard, 2013). Fenwick (2015a) posits that humans and the technologies they use are not separate entities: "material and social forces are interpenetrated in ways that have important implications for how we might examine their mutual constitution in educational processes and events" (p. 14). Not only is there an interaction between humans and materials such as technology but also a symbiotic relationship.

New technologies have moved us from an era of scarcity of information to an era of abundance (Weller, 2011). Social media now make it possible to communicate across networks on a global scale, outside the traditional classroom bound by brick walls. Communication on such a global scale would have been unimaginable not long ago. Data and data storage have evolved under the influence of emerging technologies. Instead of capturing data and storing it in a database, we now deal with large data-streams stored in the cloud, which might be represented and visualized using algorithms and machine learning. This presents interesting opportunities to learn from data, revealing with it hidden insights but important challenges as well.

Questions have been raised about how stakeholders – learners, educators, and administrators – in the educational process might manage and access all these levels of information and communication effectively. Computer scientists have suggested opportunities for automated data filtering and analysis that could do exactly that: sift through all data available and provide learners with connections to and recommendations for their preferred information, people, and tools, and in doing so personalize the learning experience and aid learners in the management and deepening of their learning (Siemens, Dawson, & Lynch, 2013). In addition, examples of research using huge institutional datasets are emerging, made possible by accessing data from traces left behind by learner activity (Xu & Smith Jaggars, 2013).

In discussing changes big data might force onto professional practice, Fenwick (2015b) highlights, for instance, the "reduction of knowledge in terms of decision making. Data analytics software works from simplistic premises: that problems are technical, comprised of knowable, measurable parameters, and can be solved through technical calculation. Complexities of ethics and values, ambiguities and tensions, culture and politics and even the context in which data is collected are not accounted for" (p. 70).

This is an important issue. She further emphasizes that the current developments involving data might change our everyday practice in ways that may not quite be understood when implemented. For instance, she highlights equality issues that arise when there is a dependence on comparison and prediction (Fenwick, 2015b). Moreover, her research led to the conclusion that research methodologies taught to prospective educators and educational researchers are completely inadequate in dealing with the big datasets available to enhance their practice. Furthermore, she wonders about "the level of professional agency and accountability. Much data accumulation and calculation is automated, which opens up new questions about the autonomy of algorithms and the attribution of responsibility when bad things happen" (p. 71). These are serious questions that need careful consideration. This chapter will address some of the challenges related to educational data mining and analytics using large datasets in research and user data in algorithms for learner support. It will also explore the impact of automation and the possible dehumanizing effects of replacing human communication and engagement in learning with technology.

THE OPPORTUNITIES OF EDUCATIONAL DATA MINING IN LEARNING MANAGEMENT

Reliability and Validity

Educational data mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students and the settings in which they learn (Ed Tech Review, 2016). Educational data mining is wider than its name would imply and it goes beyond the scope of simply mining educational data for information retrieval and building a better understanding of learning mechanisms. Hence, EDM also aims at developing methods and models to predict learner behaviours using machine learning and statistical approaches governed by scientific concerns related to validity, reproducibility, and generalizability.

Learning analytics (LA) is closely related to the field of EDM and is concerned with the measurement, collection, analysis, and reporting of data about learners and their contexts for purposes of understanding and optimizing learning and the environments in which it occurs (Long & Siemens, 2011). EDM techniques and LA are being used to augment the learning process. These seem promising in aiding in the provision of effective student support, and although there is promise that these new developments might enhance education and learning, major challenges have also been identified.

To some extent, EDM is not only a research area, thriving due to the prolific contributions of researchers from all around the world, but also a science. Recently, Britain's Science Council defined science as "the pursuit of knowledge and understanding of the natural and social world following a systematic methodology based on evidence" (British Science Council, 2009). Evidence is a requirement to any claim made in the field; as in any other scientific domain, educational data mining and analytics researchers require evidence to support or reject claims and discoveries drawn from or validated by educational data.

However, a common definition of what makes good or poor evidence is not that obvious in the EDM and LA research community, which has brought together scientists from "hard" (Computer Science) and "soft" sciences (Education). We will provide here some examples of inconsistencies and procedural flaws that we have come across during our own research. Thanks to data sharing, Long and Aleven (2014) were able to contradict learning claims of a gamified approach in an intelligent tutoring system. However, sometimes sharing datasets is not sufficient; some research work requires extensive preprocessing as several choices (biases) are made during those steps that may be hard to define clearly in a research paper. Another team trying to prepare the dataset following the same rules, therefore, might not manage to do so. The nature of the software used in preprocessing can also have an impact. The implementation of key methods can vary when using R, SPSS, Matlab, and other tools, leading to potentially different conclusions.

Another contentious aspect is the *a priori* assumption or "ground truth" that a statistical model would be built upon. This is the case, for example, in competency frameworks where mapping between items and skills defined by human experts is questionable (Durand, Belacel, & Goutte, 2015). Skills, like many other latent traits, are sometimes hard to characterize. To that end, PSLC Data Shop offers an incredible environment for learning experts to test their competency frameworks, as the observed results obtained by students help them to improve and share their mapping. It is also a great tool to share datasets among EDM and LA practitioners since sharing datasets is valuable in identifying problems. Sharing should become commonplace and no major published results should be seriously considered without the possibility for other teams to validate the claims.

Some other issues might raise questions, such as considering statistical studies and particularly linear correlational measures. EDM gathers researchers with different practices and with different perspectives on what could or should not be used as evidence. While in "hard science" a significant Pearson correlation below r=.5 would be systematically considered weak, it is usual in "soft science" to consider r=.3 values to be strong, especially regarding personality traits. Psychologists even call this .3 threshold, the "personality coefficient" because most relationships between personality traits and behaviours tend to be around that value, including the relationship between competency and performance (Mischel, 1968, p. 78). Work done in EDM regarding sentiment analysis (Wen, Yang, & Rosé, 2014) provides such an example, where it is difficult to provide computational outcomes from meaningful "soft" science research results on dropout rates in MOOCs. It might also be suggested that the topic under investigation would be better researched through qualitative techniques. However, relationships in this form of quantitative EDM remain weak when the intent is to infer predictions. Specifically, a .3 correlation that by definition explains 9% of the variance in the criterion may be of limited value in predictions in the area of sentiment analysis.

Several statistics tests can be significant as well but not really truthful regarding the accuracy of the results. El Emam (1998) evaluated how the Chi-Square test could be misleading in evaluating the predictive validity of a classifier, showing that same Chi-Square results could prove either strong or weak accuracies. More recently, Gonzalez-Brenes and Huang (2015) proposed the Leopard metric as a standard way of evaluating adaptive tutoring systems and increasing the evaluation results of the predictive accuracy of the system by evaluating their usefulness. They proposed to evaluate the amount of effort required by learners in those systems to achieve learning outcomes. After all, usefulness measures might be what the people using the systems are most interested in.

To that end, Ryan Baker, one of the most prominent researchers of the EDM community, in his MOOC entitled "Big Data in Education" provides examples and good practical advice to help researchers understand and check more wisely the validity of their models (generalizability, ecological, construct, and predictive, substantive, and content validity). In his course, Baker emphasized using Kappa and even better A' to measure respectively how a "detector is better than chance" and "the probability a detector will correctly identify" a specific trait to overcome some of the flaws of other metrics for classifiers like accuracy, ROC, precision, and recall or Chi-Square (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014, p. 492). However, A' and Kappa use seems limited in EDM publications so far.

Critically, we would like to emphasize the importance of research integrity. It might be appealing in EDM and LA to provide "made up" results. Our own work has shown that obtaining tangible results usually requires many attempts, much work is done without any guaranty of success, and the validation process appears problematic. So far, no major cases of falsifying results have been revealed but providing guidelines regarding transparency should be of greater concern to avoid potential future cases of fraud and misconduct, as observed in other scientific areas (Gupta, 2013).

We argue that in the developing fields of LA and EDM, the scientific ideal is ambitious and requires that researchers carefully check the scientific robustness of their claims. Even though research is fuelled by grants based on promises that it will impact human learning in the near future, it is important to take the time to safeguard the scientific integrity of the emerging fields of LA and EDM. This requires careful consideration by us all of the methods used and the results obtained.

The Challenges of Qualitative Data Analysis

If we look at the development of educational research over the past decades, there is a distinct movement from quantitative towards qualitative research (Gergen, Josselson, & Freeman, 2015). Psychologists increasingly support the idea that the intricacies of learning and knowing cannot be determined by testing individuals' behaviour alone. The study of the richness of learner actions and thinking in relation to the society they live in, and their communications with those in their knowledge networks, provides a much deeper, more inclusive, and critically cultural understanding of people's knowledge development and learning (Christopher, Wendt, Marecek, & Goodman, 2014; Denzin & Lincoln, 2011; Gergen et al., 2015).

The current technology-rich learning environment is not just a walled-in classroom but involves global network communications too; these encompass reflexive narrative and rich imagery, challenging researchers to re-invent their research methodologies. Moving beyond end-of-course surveys to reveal student satisfaction, mining the data produced by learners, analyzing the narrative, images, and visualizations produced during the online learning experience — all of these offer options for understanding the rich tapestry of learning interactions. Analyzing the fundamental dimensions in the changing assemblages of words and images on social media that now form part of the learning environment might get more to the heart of the learning process than official course evaluations.

Research on PLENK2010 and CLOM REL 2014, two massive open online courses (MOOCs), has highlighted the challenges that such research involves (Fournier & Kop, 2015; Kop, Fournier, & Durand, 2014). Previous MOOC research provided both bigger and richer datasets than ever before, with powerful tools to visualize patterns in the data, especially on digital social networks. The work of uncovering such patterns, however, provided more questions than answers from the pedagogical and technical contexts in which the data were generated. Moving towards a qualitative approach in trying to understand why MOOC participants produced the data that they did prompted a critical reflection on what big data, EDM, and LA could and could *not* tell us about complex learning processes and experiences.

Boyd (2010) expressed it in the following way:

> Much of the enthusiasm surrounding Big Data stems from the opportunity of having easy access to massive amounts of data with the click of a finger. Or, in Vint Cerf's words, "We never, ever in the history of mankind have had access to so much information so quickly and so easily." Unfortunately, what gets lost in this excitement is a critical analysis of what this data is and what it means. (p. 2)

In dealing with so much data and information so quickly, researchers need to envisage the optimal processes and techniques for translating data into understandable, consumable, or actionable modes of representation in order for results to be useful and accessible for audiences to digest. The ability to communicate complex ideas effectively is critical in producing something of value that translates research findings into practice. Questions have been raised about how stakeholders in the educational process (i.e., learners, educators, and administrators) might access, manage, and make sense of all these levels of information effectively; EDM and LA methods hint at how automated data filtering and analysis could do exactly that. This can lead to potentially rich inferences about learning and learners but also raise many new interesting research questions and challenges in the process. In so doing, researchers must strive to demonstrate how the data are meaningful, as well as appealing to various stakeholders in the educational process while engaging in responsible innovation with thoughtful research designs and implementations (Berland, Baker, & Blikstein, 2014).

Algorithms, Serendipity, and the "Human" in Learning: A Critical Look at Learning Analytics

A body of literature is slowly developing in EDM and LA. Essentially, it is not easy to use technology to analyze learning or use predictive analytics to advance learning. Issues around the development of algorithms and other data-driven systems in education lead to questions about what these systems actually replace and whether this replacement is positive or negative. Secondly, who influences the content of data-driven systems and what value might they add to the educational process?

In online education, but also in a connectivist networked environment (Jones, Dirckinck-Holmfeld, & Lindström, 2006), communication and dialogue between participants in the learning endeavor have been at the heart of a quality learning experience. This human touch is a necessary component in developing learning systems and environments (Bates, 2014). The presence and engagement of knowledgeable others has always been seen as vital to extend the ideas, creativity, and thinking of participants in formal learning settings, but also in online networks of interest (Jones et al., 2006).

When developing data-driven technologies for learning, it seems important to harness this human element somehow for the good of the learning process. This means that in the filtering of information, or the asking of Socratic questions, the aggregation of information should be mediated via human beings (Kop, 2012). Social microblogging sites such as Twitter have been shown to do this successfully, as "followers," who provide information and links to resources, have been chosen by the user and are seen to be valuable and credible (Bista, 2014; Kop, 2012; Stewart, 2015). In algorithms, these judgements are difficult to achieve, but perhaps a combination of recommender systems, based on data, and support and scaffolding applications based on communication, would facilitate this.

It is important to consider who influences the content of data-driven systems and what value they might add to the educational process. Furthermore, not only do the affordances and effectiveness of new technologies need to be considered, but also a reflection on the ethics of moving from a learning environment characterized by human communication to an environment that includes technical elements over which the learner has little or no control.

One of the problems highlighted in the development of algorithms is the introduction of researcher biases in the tool, which could affect the quality of the recommendation or search result (Hardt, 2014). Proper training of the people working with the data can make all the difference (Fenwick, 2015b; Boyd & Crawford, 2012). Presently, computer scientists and mathematicians, who do not necessarily have a background in the social sciences, produce the applications. As Boyd and Crawford (2012) compellingly argue:

> When computational skills are positioned as the most valuable, questions emerge over who is advantaged and who is disadvantaged in such a context. This, in its own way, sets up new hierarchies around "who can read the numbers," rather than recognizing that computer scientists and social scientists both have valuable perspectives to offer. Significantly, this is also a gendered division. Most researchers who have computational skills at the present moment are male and, as feminist historians and philosophers of science have demonstrated, who is asking the questions determines which questions are asked. (p. 674)

Boyd and Crawford (2012) suggest that computer scientists and social scientists should work together to develop bias-free, high quality analytics tools, and that teamwork with people in different fields might also be fruitful for the mining and analysis of big data. Of course, the expansion and availability of data has also made it attractive to make use of them, but there are again some challenges. Human beings for the most part get their information from sources that they trust, but as Fenwick (2015b) suggests, the use of new techniques might change "everyday practice and responsibilities in ways that may not be fully recognised" (p. 71). She highlights, for instance, that a reliance on comparison and prediction "can be self-reinforcing and reproductive, augmenting path dependency and entrenching existing inequities," especially if the people producing the algorithms are not aware of the reinforcement of stereotypes when big data is not used carefully.

Furthermore, we should not underestimate the fact that most of the algorithms currently in use were produced for economic gain and not to enhance deeper levels of learning or add value to society. As argued by Kitchin (2015), "Software is not simply lines of code that perform a set of instructions, but rather needs to be understood as a social product that emerges in contingent, relational and contextual ways, the outcome of many minds situated with diverse social, political and economic relations" (p. 5). Clearly, the development of automated algorithm systems has another inherent problem wherein it might be hard to point a finger towards who is responsible when things go wrong.

Some Ethical Considerations

Open learning environments combined with powerful data analysis tools and methods bring new affordances and support for learning but also highlight important ethical issues and challenges that move learners from an environment characterized by human communication to one that includes technical elements over which the learner has little or no control. Much of the commercial effort in Web development is informed by big data and is lacking in any innovative educational insights (Atkinson, 2015). We agree that "It is the scholarship and research informed learning design itself, grounded in meaningful pedagogical and andragogical theories of learning that will ensure that technology solutions deliver significant and sustainable benefits" to education (Atkinson, 2015, p. 7).

The dynamic pace of technological innovation, including EDM and LA, also requires the safeguarding of privacy in a proactive manner. In order to achieve this goal, researchers and system designers in the fields of EDM and advanced analytics must practice responsible innovation that integrates privacy-enhancing technologies directly into their products and processes (Cavoukian & Jonas, 2012). According to Oblinger (2012), "Analytics is a matter of culture – a culture of inquiry: asking questions, looking for supporting data, being honest about strengths and weaknesses that the data reveals, creating solutions, and then adapting as the results of those efforts come to fruition" (p. 98).

With this in mind, we strongly recommend that those designing and building next generation analytics ensure that they are informed by Privacy by Design. This entails mindfulness and responsible practice involving accountability, research integrity, data protection, privacy, and consent (Cavoukian & Jonas, 2012; Cormack, 2015). The line between private and public data is increasingly becoming blurred as more opportunities to participate in open learning environments are created and as data about participants, their activities, their interactions, and their behaviours are made accessible through social media, such as Facebook, Twitter, Google, and potentially any other social media tool available online. In the context of big data, we agree with the European Data Protection Supervisor (2015) who states that "People want to understand how algorithms can create correlations and assumptions about them, and how their combined personal information can turn into intrusive predications about their behaviour" (p. 10).

CONCLUSION

Significant questions about truth, control, transparency, and power in big data studies also need to be addressed. Pardo and Siemens (2014) maintain that keeping too much data (including student digital data, privacy-sensitive data) for too long may actually be harmful and lead to mistrust of the system or institution that has been entrusted to protect personal data. Discussions around big data ethics have underscored important methodological concerns related to data cleaning, data selection and interpretation (Boyd & Crawford, 2012), the invasive potential of data analytics, as well as the potential dehumanizing effects of replacing human communication and engagement with automated machine-learning algorithms and feedback. Researchers and developers must be mindful of the affordances and limitations of big data (including data mining and predictive learning analytics) in order to construct useful future directions (Fenwick, 2015b). Researchers should also work together in teams to avoid some of the inherent fallacies and biases in their work, and to tackle the important issues and challenges in big data and data-driven systems in order to add value to the educational process.

REFERENCES

Atkinson, S. P. (2015). Adaptive learning and learning analytics: A new learning design paradigm. BPP Working paper. https://spatkinson.files.wordpress.com/2015/05/atkinson-adaptive-learning-and-learning-analytics.pdf

Bates, T. (2014). Two design models for online collaborative learning: Same or different? http://www.tonybates.ca/2014/11/28/two-design-models-for-online-collaborative-learning-same-or-different/

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19, 206–220.

Bista, K. (2014). Is Twitter an effective pedagogical tool in higher education? Perspectives of education graduate students. *Journal of the Scholarship of Teaching and Learning*, 15(2), 83–102. http://files.eric.ed.gov/fulltext/EJ1059422.pdf

Bouchard, P. (2013). Education without a distance: Networked learning. In T. Nesbit, S. M. Brigham, & N. Taber (Eds.), *Building on critical traditions: Adult education and learning in Canada*. Toronto: Thompson Educational Publishing.

Boyd, D. (2010). Privacy and publicity in the context of big data. Paper presented at the 19th International Conference on World Wide Web (WWW2010), 29 April 2010, Raleigh, North Carolina, USA. http://www.danah.org/papers/talks/2010/WWW2010.html

Boyd, D., & Crawford, K. (2012). Critical equations for Big Data. *Information, Communication & Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878

British Science Council. (2009). What is science? http://www.sciencecouncil.org/definition

Cavoukian, A., & Jonas, J. (2012). Privacy by design in the age of big data. https://privacybydesign.ca/content/uploads/2012/06/pbd-big_data.pdf

Cormack, A. (2015). A data protection framework for learning analytics. Community.jisc.ac.uk. http://bit.ly/1OdIIKZ

Christopher, J. C., Wendt, D. C., Marecek, J., & Goodman, D. M. (2014) Critical cultural awareness: Contributions to a globalizing psychology. *American Psychologist*, 69, 645–655. http://dx.doi.org/10.1037/a0036851

Denzin, N., & Lincoln, Y. (Eds.) (2011). The Sage handbook of qualitative research (4th ed.). Thousand Oaks, CA: Sage.

Durand, G., Belacel, N., & Goutte, C. (2015) Evaluation of expert-based Q-matrices predictive quality in matrix factorization models. *Proceedings of the 10th European Conference on Technology Enhanced Learning* (EC-TEL '15), 15–17 September 2015, Toledo, Spain (pp. 56–69). Springer. doi:10.1007%2F978-3-319-24258-3_5

Ed Tech Review (2016). Educational Data Mining (EDM). http://edtechreview.in/dictionary/394-what-is-educational-data-mining

El Emam, K. (1998). The predictive validity criterion for evaluating binary classifiers. *Proceedings of the 5th International Software Metrics Symposium* (Metrics 1998), 20–21 November 1998, Bethesda, MD, USA (pp. 235–244). IEEE Computer Society. http://ieeexplore.ieee.org/document/731250/

European Data Protection Supervisor. (2015). Leading by example: The EDPS strategy 2015–2019. https://secure.edps.europa.eu/EDPSWEB/edps/site/mySite/Strategy2015

Fenwick, T. (2015a). Things that matter in education. In B. Williamson (Ed.), *Coding/learning, software and digital data in education*. University of Stirling, UK. http://bit.ly/1NdHVbw

Fenwick, T. (2015b). Professional responsibility in a future of data analytics. In B. Williamson (Ed.) *Coding/learning, software and digital data in education*. University of Stirling, UK. http://bit.ly/1NdHVbw

Fournier, H., & Kop, R. (2015). MOOC learning experience design: Issues and challenges. *International Journal on E-Learning*, 14(3), 289–304.

Gergen, J. K., Josselson, R., & Freeman, M. (2015). The promise of qualitative inquiry. *American Psychologist*, 70(1), 1–9.

Gonzalez-Brenes, J., & Huang, Y. (2015). Your model is predictive — but is it useful? Theoretical and empirical considerations of a new paradigm for adaptive tutoring evaluation. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Education Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 187–194). International Educational Data Mining Society.

Gupta, A. (2013). Fraud and misconduct in clinical research: A concern. *Perspectives in Clinical Research.* 4(2), 144–147. doi:10.4103/2229-3485.111800.

Hardt, M. (2014). How big data is unfair: Understanding sources of unfairness in data driven decision making. https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de

Jones, C., Dirckinck-Holmfeld, L., & Lindström, B. (2006). A relational, indirect, meso-level approach to CSCL design in the next decade. *International Journal of Computer Supported Collaborative Learning*, 1(1), 35–56.

Kansanen, P., & Meri, M. (1999). The didactic relation in the teaching–studying–learning process. *TNTEE Publications*, 2, 107–116. http://www.helsinki.fi/~pkansane/Kansanen_Meri.pdf

Kitchin, R. (2015). Foreword: Education in code/space. In B. Williamson (Ed.), *Coding/learning, software and digital data in education*. University of Stirling, UK. http://bit.ly/1NdHVbw

Kop, R. (2012). The unexpected connection: Serendipity and human mediation in networked learning. *Educational Technology & Society*, 15(2), 2–11.

Kop, R., Fournier, H., & Durand, G. (2014). Challenges to research in massive open online courses. *Merlot Journal of Online Learning and Teaching*, 10(1). http://jolt.merlot.org/vol10no1/fournier_0314.pdf

Long, Y., & Aleven, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems* (ITS 2014), 5–9 June 2014, Honolulu, HI, USA (pp. 378–387). New York: Springer. doi:10.1007/978-3-319-07221-0_47

Long, C., & Siemens, G. (2011, September 12). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5). http://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Meyer, J. H. F., & Land, R. (2006). *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge.* London: Routledge.

Mischel, W. (1968). *Personality and assessment*. London: Wiley.

Oblinger, D. G. (2012, November 1). Analytics: What we're hearing. *EDUCAUSE Review*. http://er.educause.edu/articles/2012/11/analytics-what-were-hearing

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology*, 45(3), 487–501.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450.

Siemens, G., Dawson, D., & Lynch, L. (2013). *Improving the quality and productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics*. SoLAR. https://www.itl.usyd.edu.au/projects/SoLAR_Report_2014.pdf

Stewart, B. (2015). Open to influence: What counts as academic influence in scholarly networked Twitter participation. *Learning, Media and Technology*, 40(3), 287–309. doi:10.1080/17439884.2015.1015547

Weller, M. (2011). A pedagogy of abundance. *Spanish Journal of Pedagogy*, 249, 223–236.

Wen, M., Yang, D., & Rosé, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 130–137). International Educational Data Mining Society.

Xu, D., & Smith Jaggars, S. (2013). Adaptability to online learning: Differences across types of students and academic subject areas. CCRC Working Paper No. 54. Community College Research Center, Teachers College, Columbia University. http://anitacrawley.net/Reports/adaptability-to-online-learning.pdf

Chapter 28: Unpacking Student Privacy

Elana Zeide

Center for Information Technology Policy, Princeton University, USA
Information Society Project, Yale Law School, USA
Information Law Institute, New York University School of Law, USA

DOI: 10.18608/hla17.028

ABSTRACT

The learning analytics and education data mining discussed in this handbook hold great promise. At the same time, they raise important concerns about security, privacy, and the broader consequences of big data-driven education. This chapter describes the regulatory framework governing student data, its neglect of learning analytics and educational data mining, and proactive approaches to privacy. It is less about conveying specific rules and more about relevant concerns and solutions. Traditional student privacy law focuses on ensuring that parents or schools approve disclosure of student information. They are designed, however, to apply to paper "education records," not "student data." As a result, they no longer provide meaningful oversight. The primary federal student privacy statute does not even impose direct consequences for noncompliance or cover "learner" data collected directly from students. Newer privacy protections are uncoordinated, often prohibiting specific practices to disastrous effect or trying to limit "commercial" use. These also neglect the nuanced ethical issues that exist even when big data serves educational purposes. I propose a proactive approach that goes beyond mere compliance and includes explicitly considering broader consequences and ethics, putting explicit review protocols in place, providing meaningful transparency, and ensuring algorithmic accountability.

Keywords: MOOCs, virtual learning environments, personalized learning, student privacy, education data, ed tech, FERPA, SOPIPA, data ethics, research ethics

First, a caveat: the descriptions in this chapter should not be used as a guide to compliance, since legal requirements are constantly changing. It instead provides ways to think about the issues that people discuss under the banner of "student privacy" and the broader issues often neglected. In the United States, privacy rules vary across sectors. Traditional approaches to student privacy, most notably the Family Educational Rights and Privacy Act (FERPA),[1] rely on regulating how schools share and allow access to personally identifiable student information maintained in education records. They use informed consent and institutional oversight over data disclosure as a means to ensure that only actors with legitimate educational interests can access personally identifiable student information. This approach aligns with the Fair Information Practice Principles – typically notice, choice, access, and right to amend – that have been at the core of most privacy regulation since the early 1970s. These early privacy rules also focus primarily on disclosure of student information without addressing educators' collection, use, or retention of education records.

Newer approaches to student privacy tend to simply prohibit certain practices or require them to serve "educational" purposes. Blunt prohibitions are often crafted too crudely to work within the existing education data ecosystem, let alone support growth and innovation. "Education" purpose restrictions may limit explicit "commercial" use of student data, but they do not deal with the more nuanced issues raised by learning analytics and educational data mining even when used by educators for educational purposes. They do not consider the ways that using big data to serve education may not serve the interests of all educational stakeholders. It is difficult for categorically prohibitive legislation to be sufficiently flexible to match

[1] Family Educational Rights and Privacy Act (2014): see https://www2.ed.gov/policy/gen/guid/fpco/ferpa/index.html?src=rn and https://www2.ed.gov/policy/gen/guid/fpco/pdf/ferparegs.pdf

the fast pace of technological change and the highly contextualized decision making in learning spaces. Data scientists and decision makers using learning analytics and education data mining must go beyond mere compliance through deliberate foresight, transparency, and accountability to ensure that data-driven tools achieve their goals, benefit the education system, and promote equity in broader society.

EDUCATION RECORD PRIVACY

The first wave of student privacy panic occurred in the late 1960s and early 1970s. Schools began to collect a wider array of information about students. Educators and administrators routinely shared student information on an ad hoc and often undocumented basis (Divoky, 1974).

FERPA's Default against School Disclosure

In response, Congress passed the primary federal statute governing student data, FERPA, in 1974. FERPA gives three rights to parents and "eligible students" over 18 or enrolled in postsecondary education ("parents," as shorthand). Federally funded schools, districts, and state education agencies must provide parents with access to education records maintained by the education institution or agency ("education actors," as shorthand) and the ability to challenge their accuracy. Education actors must also get parents' permission before sharing personally identifiable student information, subject to many exceptions that allow schools to consent on their behalf.[2]

FERPA focuses on limiting the disclosure of personally identifiable student information by educational institutions and agencies to approved recipients with legitimate educational interests. To meet FERPA's requirements, schools must obtain parents' written consent before sharing personally identifiable information maintained in a student's educational record unless one of several exceptions applies. In practice, the exceptions swallow the rule, and educators, not parents or students, make most privacy decisions (Zeide, 2016a).

Schools Authorizing Disclosure to Serve "Educational" Interests

The school official exception delegates the bulk of data-related decision making to schools and districts. Schools can share student personally identifiable student information without prior consent if the recipient is 1) performing services on their behalf and 2) has a "legitimate educational interest" in accessing such information; and, ostensibly, 3) has taken reasonable measures to exercise direct control over the information.[3] Educators decide what qualifies someone to be a school official and what constitutes a legitimate educational interest, but do not have to define these terms in any substantive detail (US Department of Education, n.d.). As a result, most rely on criteria so broad as to encompass almost any circumstance (Zeide, 2016a). Schools rarely take active measures to control recipients' detailed information practices, relying instead on terms of service or contracts between the parties as the means of "direct control" (Reidenberg et al., 2013).

Researchers Barred from Repurposing Student Data

FERPA places more stringent requirements on how educational actors share information with researchers. Under the studies exception, they must do so pursuant to a written contract with specific terms. Studies must be for the purpose of "developing, validating, or administering predictive tests; administering student aid programs; or improving instruction."[4] Researchers may only use personally identifiable student information for specified purposes and destroy the data once it is no longer needed.

Compliance-Oriented Enforcement

Educational actors have no direct accountability for FERPA violations. The statute is about putting a structure into place rather than preventing specific privacy violations. As a result, it does not impose consequences for individual instances of noncompliance. Students and educators cannot sue for violations under the statute (US Supreme Court, 2002). Instead, the US Department of Education (ED) has the power to withdraw all federal funding, including support in the form of federal student loans, if an educational institution or agency has a "policy or practice" of noncompliance.[5] However, the Department has never taken this dramatic action since the statute's enactment over forty years ago (Zeide, 2016b; Daggett, 2008). Since such a drastic measure would hurt the very students FERPA seeks to protect, the agency instead focuses on bringing education institutions into compliance. It is unlikely ED will ever pursue such a "nuclear" option (Solove, 2012).

STUDENT DATA PRIVACY

For almost forty years, stakeholders predominantly accepted FERPA's protection as sufficient despite minimal transparency, individual control over information, or consequences for specific violations in practice. FERPA's regulatory mechanisms no longer provide

[2] Family Education Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g; 34 CFR Part 99 (2014); 34 CFR § 99.31 (exceptions): https://www.law.cornell.edu/uscode/text/20/1232g; https://www2.ed.gov/policy/gen/guid/fpco/pdf/ferparegs.pdf.

[3] Id. § 99.31(a)(1) (School Official Exception).

[4] Id. § 99.31(a)(6) (Studies Exception).

[5] 20 U.S.C. § 1232g(b)(1)–(2) (Policies or Practice Provision).

sufficient reassurance for stakeholders, in part because it regulates education records, not student data. The statute provides only narrow protection in terms of the information it covers, the actions it relates to, and the entities to which it applies in an age of big data.

From Education Records to Student Data

Low-cost storage, instantaneous transfer over connected networks, and cloud-based servers create an unprecedented volume, velocity, and variety of "big data" (Mayer-Schönberger & Cukier, 2014). Student information no longer means paper "education records" locked away in school filing cabinets, but rather interoperable, instantly transferable data stored on cloud servers. Interactive educational tools and platforms generate more information about students with more detail than has previously been possible. Data-mined information from out-of-classroom sources, like school ID geolocation and social media, goes far beyond traditional expectations regarding education records (Alamuddin, Brown, & Kurzweil, 2016). Even when mined student information is publicly available, many stakeholders find the notion of systematic collection and analysis of student data unsettling (Watters, 2015). The automatic capture of clickstream-level data about students, the permeability of cloud computing networks, and the infinite utility of big data prompts new privacy concerns (Singer, 2013).

FERPA's reliance on parental, student, or school oversight of recipients' information practices may not be possible, let alone practical or meaningful, given the quantity and complexity of big data and the automated transmission of information in interactive, digitally mediated environments. The statute does not even address schools' own privacy practices or cover new independent education providers, like massive open online courses (MOOCs), which collect information directly from users in "learning environments" but receive no federal funding. Stakeholders have little idea about what information schools and companies collect on students and how they use them (Barnes, 2014). They can't be sure that educators and data recipients even adhere to the privacy promises they make – especially when FERPA imposes no direct accountability for non-compliance.

Proliferation of New Student Privacy Protections

Since 2013, state policymakers responded to stakeholder panic by introducing over 410 student privacy bills: 36 states have passed 73 of these into law. On the federal level, legislators proposed amendments to FERPA and bills that would directly regulate the companies and organizations receiving student information. The vast majority of protective measures apply to federally funded P–12 public schools, but there is no consensus about what concerns matter and what "student privacy" means. This is clear from the incredible variety of ways districts, researchers, institutions, companies, states, and federal policymakers propose to protect student data (Center for Democracy and Technology, 2016; DQC, 2016; Vance, 2016).

Almost all reform measures reflect the need for more transparency, accountability, and baseline data safety, security, and governance protocols. Many simply continue FERPA's focus on how schools share information with third-party vendors and education researchers. Several explicitly prohibit school collection of certain types of information or from outside sources like social media. Some measures regulate data-reliant service providers directly (Center for Democracy and Technology, 2016; DQC, 2016; Vance, 2016).

Self-Regulation Supplements

More flexible approaches to privacy governance involve self-regulation. Over 300 companies have signed a Student Privacy Pledge,[6] created by the Future of Privacy Forum and the Software & Information Industry Association, which includes ten principles such as not selling student data. Signatories risk FTC enforcement if they do not abide by their promises (Singer, 2015). The US Department of Education, education organizations, and privacy experts are continuously releasing new best practice guidelines and privacy toolkits (Krueger, 2014; Privacy Technical Assistance Center, 2014). For stakeholders to have sufficient trust in these rules, however, there must be sufficient transparency about information practices, consideration regarding learning analytics purposes and potential outcomes, and accountability for noncompliance.

STUDENT PRIVACY GAPS

While the latest round of student privacy regulation has prompted much more explicit governance of student data and some sorely needed transparency, most reform measures still suffer from many of FERPA's flaws. Most students and stakeholders still have no concrete sense of what information is contained in education records, vague notions of how data can be used to their benefit, and minimal reassurance about what protections are in place (Prinsloo & Rowe, 2015; Rubel & Jones, 2016; Zeide, 2016a).

Minimal Meaningful Consent and Oversight

FERPA and similar rules rely on parental or school oversight of disclosure as a way to ensure that only appropriate recipients can access student data. This may not be possible, let alone practical or meaningful, given the quantity and complexity of big data and the automated transmission of information in interactive,

[6] http://studentprivacypledge.org/?page_id=45

digitally mediated environments. Contractual provisions create some more structure, but require schools to monitor third-party information practices and bring expensive lawsuits for enforcement. Finally, limiting disclosure doesn't work as well to prevent inappropriate use of student information since recipients who initially use this data for "legitimate education interests" often serve corporate or research interests at the same time (Young, 2015; Zeide, 2016a).

Crude Categorical Prohibitions

Some regulations attempt to address this problem by completely barring specific data collection, use, and repurposing. This often leads to problematic outcomes that conflict with current data use in the education system and unnecessarily restricts promising learning analytics and educational data mining. In Florida, for example, a ban on collecting biometric information conflicted with existing practices and legal obligations regarding special education students. As a result, many states have had to suspend or amend their initial attempts at ensuring student privacy. Erasure rules severely limit the potential for longitudinal studies and frequently often conflict with other record-keeping obligations imposed by state law (Vance, 2016).

Limits of Education Purpose Limitations

Many new laws follow the model of California's Student Online Personal Information Protection Act (SOPIPA), which covers entities providing education (K–12) services. They regulate how these actors use student information directly, rather than trying to do so through school oversight. Online providers of such services must have contracts with schools, erase student information upon request, and cannot create learner profiles that don't serve specified "K–12 purposes."[7]

Regulations that limit student data to "educational" use or purposes attempt to prevent commercial misuse by for-profit entities. Purpose limitations, however, do not address more nuanced issues raised by learning analytics and education data mining. Purpose limitation rules rest on the assumption of a consensus about what constitutes an "educational" purpose. They do not consider ways that institutions or researchers might prioritize goals other than the immediate educational interests of learner data subjects while still legitimately using data to manage institutional resources, improve the education system, or shed insight on learning science. Schools might, for example, use predictive data to exclude, rather than encourage, marginal students to save resources or improve rankings (Ashman et al., 2014; Drachsler & Greller, 2016; Rubel & Jones, 2016; Selwyn, 2014; Slade, 2016).

Leaving Out Learner Data

Most new laws do not address information held in higher education institutions. They do not address "learner" information collected by virtual learning environments independent of traditional, federally funded education institutions. Instead, the more permissive commercial privacy regime governs data collected and used by these private entities (in the absence of applicable state law). This means that use and disclosure of this "learner" data is limited by consumer privacy policies, which are notorious for being incomprehensible, overly broad, and open to change without notice (Jones & Regner, 2015; Polonetsky & Tene, 2014; Young, 2015; Zeide, 2016a).

EDUCATION DATA ETHICS

Society grapples with these issues across sectors, but they are particularly acute in education environments. As individuals who seek to improve education experiences and operate with integrity, it is easy to lose sight of how revolutionary the information practices involved in learning analytics and education data mining are compared to traditional education information practices and norms about student data. Students rarely have a realistic choice to opt out of mainstream data-driven technologies. Education data subjects are more vulnerable than those in typical consumer contexts, not only because they might be children, but also because learning requires some degree of risk-taking for intellectual growth. There are still unresolved issues about whether these tools may inadvertently reduce, rather than expand equitable opportunities, undermine the broader goals of the education system, and give students less agency and make them more, not less, vulnerable (Prinsloo & Slade, 2016; Siemens, 2014).

Equitable Outcomes

It is important for those working with student data to consider how consequences may play out in an inevitably flawed reality rather that the neutral space of theoretical and technological models. Algorithmic models may inadvertently discriminate against minorities or students of lower socioeconomic status. They may have disparate impacts. Tools that predict student success could repeat past inequities instead of promoting more achievement and upward mobility. Ostensibly neutral policies can create deeply inequitable outcomes due to uneven implementation (boyd & Crawford, 2011; Citron & Pasquale, 2014; Barocas & Selbst, 2014).

[7] The statute does, however, include a carve out indicating that it "does not limit the ability of an operator to use information, including covered information, for adaptive or personalized student learning purposes." § 22584(l). As of the writing of this chapter, it is not clear how these rules will work in practice. Student Online Personal Information Protection Act (SOPIPA), CAL. BUS. & PROF. CODE §§ 22584–22585 (2014), https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201320140SB1177.

Broader Education Effects

Continuously collecting detailed information in classrooms, from cameras, or from sensors can have broader consequences. Ubiquitous surveillance and embedded assessment may have a chilling effect on student participation and expression (Boninger & Molnar, 2016; Vance & Tucker, 2016). While these practices reduce reliance on periodic high-stakes tests, they also put every moment of the learning process under scrutiny. This may ultimately undermine trust in data-driven education tools and practitioners, chilling the intellectual risk-taking required in learning environments.

Inadvertently Shifting Authority

Learning analytics and educational data mining changes not only how, but who makes pedagogical and academic decisions. Traditionally, the individuals who evaluated and made decisions about students were close at hand and relied on personal, contextualized observation and knowledge. Parents, students, or administrators with concerns about particular outcomes could go directly to the relevant decision maker for explanation. This created transparency, and an easy avenue to seek redress, thereby providing accountability.

In adopting data-driven education tools, educators change what goes into measuring learning, what goals we seek to achieve through education, and who gets to make those decisions. Automated and algorithmic pedagogical and institutional decision-making shifts the locus of authority from a traditional, physically present human to obscure technologies or remote companies and researchers. Data-driven education changes who gets to make important decisions that shape lives and the education system overall. It does so without the shift being obvious, and, in many cases, deliberate. This shift in who can access and use data shifts power relationships as well. As security expert Bruce Schneier (2008) notes, “Who controls our data controls our lives” (paragraph 5). We must explicitly consider the handoff of authority that goes with the handoff of data.

GOING BEYOND COMPLIANCE

Under the current and emerging regulatory framework, learning analytics and education data mining practitioners and consumers will have much of that power. They will accordingly bear the responsibility of defining what student privacy means. Their decisions about technological structures, conceptual models, and learning outcomes craft the rules that apply in practice to information in learning environments. These decisions need to be made thoughtfully and deliberately. It also benefits learning analytics and educational data mining as a field by cultivating the trust required for individual participation, institutional implementation, and policymaker support for learning analytics and educational data mining overall.

I recommend going beyond mere compliance to take a more proactive approach. Ideally, this involves not only anticipating potential problems, but also putting protocols in place to determine practices if they arise and open communication with data subjects and stakeholders. Key components of proactive student privacy practices include 1) considering ethical implications; 2) creating explicit protocols for review; 3) actively communicating with data subjects and stakeholders about data practices, purposes, and protection; and 4) ensuring algorithmic accountability.

Ethical Scrutiny

Learning analytics and educational data mining projects should include deliberate, proactive consideration of potential benefits and their distribution across society and time, unintended outcomes related to learning and broader society, and ethical questions regarding experiment protocols and ultimate priorities. These reflect important considerations regarding human subject experiments promulgated in the Belmont Report in 1978 and later codified and institutionalized through Institutional Review Boards (IRBs) that must approve of academic research. However, data use only inside institutions, activities categorized as “optimization” instead of research, and company practices rarely undergo similarly explicit consideration of fundamental ethical principles.

Learning analytics and educational data mining practitioners, consortia, and supporters have promulgated ethical principles to guide information practices. These raise important issues, including the importance and difficulty of user notice and consent to how data is collected, stored, processed, and shared in learning systems, given the volume of information and complexity of algorithmic analysis. They also include more abstract notions of justice and beneficence that take into account whether experimental results serve the “greater good” (Drachsler & Greller, 2016; Open University, 2017; Pardo & Siemens, 2014; Sclater & Bailey, 2015; Slade, 2016; Asilomar, 2014).

Explicit Review

Privacy and ethical considerations should be incorporated from the first stages of technology and experimental design. At a minimum, data-driven education tools should be audited for unintended bias, disparate impact, and disproportionate distribution of risk and benefits across society. A best practice would create proactive measures to address possible, but foreseeable, problematic outcomes ahead of time. Is there a point, for example, when the discrepancy between two experimental groups is so high that researchers and educators should stop A/B testing?

Projects should have predetermined points for explicit accountability and ethical review. Many companies, for example, have begun to employ their own "consumer review boards" to take data subjects' and broader society's interests into account before moving forward on experiments and again before publication (Calo, 2013; Jackman & Kanerva, 2016; Tene & Polonetsky, 2015).

Aggressive Transparency

Ideally, learning analytics and education data mining tools and technologies should also provide meaningful transparency and algorithmic accountability. Transparency is important on both the micro- and the macro-level. Disclosing information practices helps reassure stakeholders who might panic in an absence of sufficiently specific and readily available information about learning analytics and educational data mining data practices.

Transparency and outreach about the ways that data analysis may benefit current learners – and not some future student in a land far, far, away – helps ameliorate stakeholder fears. Open and early communication also helps reduce the impression that a small elite group of scientists have tremendous control over student experiences and outcomes, and that their actions are shrouded in secrecy. It helps to recruit institutional resources to find ways to reach out to data subjects and the wider community.

Algorithmic Accountability

Transparency, however, is not enough to ensure appropriate information practices. It is a prerequisite. Documentation and accountability are also important given the stakes at issue and the obscurity of algorithmic decision making. Learners and stakeholders will want to know what evidence backs up pedagogical and institutional decision making. Ideally, learning analytics and education data mining practitioners should implement tools for algorithmic accountability. These include audits to double check that algorithmic tools perform as intended and actually promote promised outcomes.

A key piece of algorithmic accountability that will become increasingly important in affecting learners' future opportunities is the need to document algorithmic and institutional decision making to allow for due process (Diakopoulos, 2016; Kobie, 2016; Kroll et al., 2017). Learners, educators, and institutions will want to see the evidence and know about the systems that impact their academic progress and credentialing and examine the decisions that affect them (Zeide, 2016a; see also Citron & Pasquale, 2014; Crawford & Schultz, 2014). Data scientists and data-driven decision makers should be prepared to facilitate forensic examination of important decisions. Parents, for example, will want an explanation as to why their child was or was not promoted to the next grade. "Because the algorithm said so" will not be a sufficient response.

CONCLUSION

Trust is crucial to learning environments, which seek to foster intellectual experimentation and growth. As noted in a 2014 White House report on big data, "As learning itself is a process of trial and error, it is particularly important to use data in a manner that allows the benefits of those innovations, but still allows a safe space for students to explore, make mistakes, and learn without concern that there will be long term consequences for errors that are part of the learning process."

By going beyond mere compliance, those entrusted with education data can guard against potential unintended consequences of even the most well-meaning projects that might undermine the very goals they seek to achieve. The readers of this handbook entrusted with the wealth of student data should take a proactive approach that aims not at mere compliance, but goes beyond to consider broader social, ethical, and political implications. Doing so will promote trust in data-driven education and ensure that learning analytics and educational data mining achieve their revolutionary potential.

REFERENCES

Alamuddin, R., Brown, J., & Kurzweil, M. (2016). *Student data in the digital era: An overview of current practices*. Ithaka S+R. doi:10.18665/sr.283890

Ashman, H., Brailsford, T., Cristea, A. I., Sheng, Q. Z., Stewart, C., Toms, E. G., & Wade, V. (2014). The ethical and social implications of personalization technologies for e-learning. *Information & Management*, 51(6), 819–832. doi:10.1016/j.im.2014.04.003

Asilomar Convention for Learning Research in Higher Education. (2014). *Student data policy and data use messaging for consideration at Asilomar II*. Asilomar, CA. http://asilomar-highered.info/

Barnes, K. (2014, March 6). Why a "Student Privacy Bill of Rights" is desperately needed. Washington Post. http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/03/06/why-a-student-privacy-bill-of-rights-is-desperately-needed/

Barocas, S., & Selbst, A. D. (2014). *Big data's disparate impact*. SSRN Scholarly Paper. Elsevier. http://papers.ssrn.com/abstract=2477899

Boninger, F., & Molnar, A. (2016). *Learning to be watched: Surveillance culture at school.* Boulder, CO: National Education Policy Center. http://nepc.colorado.edu/files/publications/RB%20Boninger-Molnar%20Trends.pdf

boyd, d., & Crawford, K. (2011). Six provocations for big data. SSRN Scholarly Paper. Elsevier. http://papers.ssrn.com/abstract=1926431

Calo, R. (2013). Consumer subject review boards: A thought experiment. *Stanford Law Review Online*, 66, 97.

Center for Democracy and Technology (2016, October 5). *State student privacy law compendium.* https://cdt.org/insight/state-student-privacy-law-compendium/

Citron, D. K., & Pasquale, F. A. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, 89. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2376209

Crawford, K., & Schultz, J. (2014). Big data and due process: Toward a framework to redress predictive privacy harms. *Boston College Law Review*, 55(1), 93.

Daggett, L. M. (2008). FERPA in the twenty-first century: Failure to effectively regulate privacy for all students. *Catholic University Law Review*, 58, 59.

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62. doi:10.1145/2844110

Divoky, D. (1974, March 31). How secret school records can hurt your child. *Parade Magazine*, 4–5.

DQC (Data Quality Campaign) (2016). *Student data privacy legislation: A summary of 2016 state legislation.* http://2pido73em67o3eytaq1cp8au.wpengine.netdna-cdn.com/wp-content/uploads/2016/09/DQC-Legislative-summary-09232016.pdf

Drachsler, H., & Greller, W. (2016). Privacy and analytics – it's a DELICATE issue. A checklist to establish trusted learning analytics. Open Universiteit. http://dspace.ou.nl/bitstream/1820/6381/1/Privacy%20a%20DELICATE%20issue%20(Drachsler%20%26%20Greller)%20-%20submitted.pdf

Jackman, M., & Kanerva, L. (2016). Evolving the IRB: Building robust review for industry research. *Washington and Lee Law Review Online*, 72(3), 442.

Jones, M. L., & Regner, L. (2015, August 19). Users or students? Privacy in university MOOCS. *Science and Engineering Ethics*, 22(5), 1473–1496. doi:10.1007/s11948-015-9692-7

Kobie, N. (2016, January 29). Why algorithms need to be accountable. *Wired UK*. http://www.wired.co.uk/news/archive/2016-01/29/make-algorithms-accountable

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*, 165. http://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2765268

Krueger, K. R. (2014). 10 steps that protect the privacy of student data. *The Journal*, 41(6), 8.

Mayer-Schönberger, V., & Cukier, K. (2014). *Learning with big data.* Eamon Dolan/Houghton Mifflin Harcourt.

Open University. (2017). Ethical use of student data for learning analytics policy. http://www.open.ac.uk/students/charter/essential-documents/ethical-use-student-data-learning-analytics-policy

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. doi:10.1111/bjet.12152

Polonetsky, J., & Tene, O. (2014). Who is reading whom now: Privacy in education from books to MOOCs. *Vanderbilt Journal of Entertainment & Technology Law*, 17, 927.

Prinsloo, P., & Rowe, M. (2015). Ethical considerations in using student data in an era of "big data." In W. R. Kilfoil (Ed.), *Moving beyond the Hype: A Contextualised View of Learning with Technology in Higher Education* (pp. 59–64). Pretoria, South Africa: Universities South Africa.

Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182.

Privacy Technical Assistance Center. (2014) *Protecting student privacy while using online educational services: Requirements and best practice*. http://ptac.ed.gov/sites/default/files/Student%20Privacy%20and%20Online%20Educational%20Services%20%28February%202014%29.pdf

Reidenberg, J. R., Russell, N. C., Kovnot, J., Norton, T. B., Cloutier, R., & Alvarado, D. (2013). *Privacy and cloud computing in public schools*. Bronx, NY: Fordham Center on Law and Information Policy. http://ir.lawnet.fordham.edu/clip/2/?utm_source=ir.lawnet.fordham.edu%2Fclip%2F2&utm_medium=PDF&utm_campaign=PDFCoverPages

Rubel, A., & Jones, K. M. L. (2016). Student privacy in learning analytics: An information ethics perspective. *The Information Society*, 32(2), 143–159. doi:10.1080/01972243.2016.1130502

Schneier, B. (2008, May 15). Our data, ourselves. *Wired Magazine*. http://www.wired.com/2008/05/security-matters-0515/all/1/

Sclater, N., & Bailey, P. (2015). *Code of practice for learning analytics*. https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics

Selwyn, N. (2014). *Distrusting educational technology: Critical questions for changing times*. New York/London: Routledge, Taylor & Francis Group.

Siemens, G. (2014, January 13). The vulnerability of learning. http://www.elearnspace.org/blog/2014/01/13/the-vulnerability-of-learning/

Singer, N. (2015, March 5). Digital learning companies falling short of student privacy pledge. *New York Times*. http://bits.blogs.nytimes.com/2015/03/05/digital-learning-companies-falling-short-of-student-privacy-pledge/

Singer, N. (2013, December 13). Schools use web tools, and data is seen at risk. *New York Times*. http://www.nytimes.com/2013/12/13/education/schools-use-web-tools-and-data-is-seen-at-risk.html

Slade, S. (2016). Applications of student data in higher education: Issues and ethical considerations. Presented at Asilomar II: Student Data and Records in the Digital Era. Asilomar, CA.

Solove, D. J. (2012). FERPA and the cloud: Why FERPA desperately needs reform. http://www.law.nyu.edu/sites/default/files/ECM_PRO_074960.pdf

Tene, O., & Polonetsky, J. (2015). Beyond IRBs: Ethical guidelines for data research. *Beyond IRBs: Ethical review processes for big data research*. https://bigdata.fpf.org/wp-content/uploads/2015/12/Tene-Polonetsky-Beyond-IRBs-Ethical-Guidelines-for-Data-Research1.pdf

US Department of Education (n.d.). FERPA frequently asked questions: FERPA for school officials. Family Policy Compliance Office. http://familypolicy.ed.gov/faq-page/ferpa-school-officials

US Supreme Court. (2002). Gonzaga Univ. v. Doe, 536 U.S. 273. https://supreme.justia.com/cases/federal/us/536/273/case.html

Vance, A. (2016). *Policymaking on education data privacy: Lessons learned.* Education Leaders Report, 2(1). Alexandria, VA: National Association of State Boards of Education.

Vance, A., & Tucker, J. W. (2016). *School surveillance: The consequences for equity and privacy*. Education Leaders Report, 2(4). Alexandria, VA: National Association of State Boards of Education.

Watters, A. (2015, March 17). Pearson, PARCC, privacy, surveillance, & trust. Hack Education: The History of the Future of Education Technology. http://hackeducation.com/2015/03/17/pearson-spy

White House. (2014). *Big data: Seizing opportunities, preserving values*. Washington, DC: Exceutive Office of the President.

Young, E. (2015). Educational privacy in the online classroom: FERPA, MOOCs, and the big data conundrum. *Harvard Journal of Law and Technology*, 28(2), 549–593.

Zeide, E. (2016a). Student privacy principles for the age of big data: Moving beyond FERPA and FIPPs. *Drexel Law Review*, 8(2), 339.

Zeide, E. (2016b, March 16). Interview with ED Chief Privacy Officer Kathleen Styles. Washington, D.C.

Chapter 29: Linked Data for Learning Analytics: The Case of the LAK Dataset

Davide Taibi[1], Stefan Dietze[2]

[1] Istituto per le Tecnologie Didattiche, Consiglio Nazionale delle Ricerche, Italy
[2] L3S Research Center, Germany

DOI: 10.18608/hla17.029

ABSTRACT

The opportunities of learning analytics (LA) are strongly constrained by the availability and quality of appropriate data. While the interpretation of data is one of the key requirements for analyzing it, sharing and reusing data are also crucial factors for validating LA techniques and methods at scale and in a variety of contexts. Linked data (LD) principles and techniques, based on established W3C standards (e.g., RDF, SPARQL), offer an approach for facilitating both interpretability and reusability of data on the Web and as such, are a fundamental ingredient in the widespread adoption of LA in industry and academia. In this chapter, we provide an overview of the opportunities of LD in LA and educational data mining (EDM) and introduce the specific example of LD applied to the Learning Analytics and Knowledge (LAK) Dataset. The LAK dataset provides access to a near-complete corpus of scholarly works in the LA field, exposed through rigorously applying LD principles. As such, it provides a focal point for investigating central results, methods, tools, and theories of the LA community and their evolution over time.

Keywords: Linked data, LAK dataset

Learning analytics (LA)[1] and educational data mining (EDM)[2] have gained increasing popularity in recent years. However, large-scale take-up in educational practice as well as significant progress in LA research is strongly dependent on the quantity and quality of available data. Being able to interpret and understand data about learning activities, including respective knowledge about the learning domain, subjects, or skills is a prerequisite for carrying out higher-level analytics. However, data as generated through learning environments is often ambiguous and highly specific to a particular learning scenario and use case, often using proprietary terminologies or identifiers for both understanding and interpreting learning-related data within the scenario, and even more, across organizational and application boundaries.

LD principles (Bizer, Heath, & Bernes-Lee, 2009) have emerged as a de facto standard for exposing data on the Web and have the potential to improve both the quantity and quality of LA data substantially by 1) enabling *interpretation* of data and 2) Web-wide *sharing* of datasets across scenarios and institutional boundaries. Facilitated through established W3C standards such as RDF and SPARQL, LD has gained significant popularity throughout the last decade, with over 1000 datasets in the recent Linked Open Data Crawl[3] alone. LD and its offspring see widespread adoption through all sorts of entity-centric approaches, such as the use of knowledge graphs for facilitating Web search, a common practice in major search engines such as Google or Bing, or the increasing adoption of Microdata and RDFa[4] for annotating Web pages with structured facts. This also led to the emergence of a growing Web of educational data (d'Aquin, Adamou, & Dietze, 2013), substantially facilitated by the availability of shared vocabularies for educational purposes and knowledge graphs such as DBpedia[5] or Freebase[6] for enriching and disambiguating data.

We argue that LD principles can act as a fundamental facilitator for scaling up LA research (d'Aquin, Dietze,

[1] http://solaresearch.org
[2] http://www.educationaldatamining.org
[3] http://linkeddatacatalog.dws.informatik.uni-mannheim.de/state/
[4] https://www.w3.org/TR/xhtml-rdfa-primer/
[5] http://dbpedia.org
[6] http://freebase.org

Herder, Hendrik, & Taibi, 2014), as well as improving performance of LA tools and methods by enabling 1) the non-ambiguous interpretation of learning data (d'Aquin & Jay, 2013) and 2) the widespread sharing of the data used for evaluating and assessing LA methods and tools in research and educational practice. After a brief summary of LD use in education, we will introduce the successful application of LD principles in the LA context of the LAK dataset.[7] The LAK dataset represents, on the one hand, a representative example of successfully applying LD principles to facilitate research in LA and, on the other, constitutes an important resource in its own right by providing access to a near-complete corpus of LA and EDM research. This is followed by a set of examples that demonstrate the benefits of applying LD principles by showcasing how new insights can be generated from such a corpus and, at the same time, provide insights into observable trends and topics in LA and EDM.

LINKED DATA IN LEARNING AND EDUCATION

Distance teaching and openly available educational data on the Web are becoming common practices with public higher education institutions as well as private training organizations realizing the benefits of online resources. This includes data 1) about *learning resources*, ranging from dedicated educational resources to more informal knowledge resources and content, and 2) data about *learning activities*.

LD principles (Heath & Bizer, 2011) offer significant opportunities for sharing, interpreting, or enriching data about both resources and activities in learning scenarios. Essentially, LD principles rely on a common graph-based representation format, the so-called Resource Description Framework (RDF),[8] a common query language (SPARQL[9]) and most notably, the use of dereferenceable URIs to name things (i.e., entities). This last feature is a key facilitator for LD as it enables the unique identification of any entity in any dataset across the Web, and hence links data across different datasets. This facilitates, for instance, an entity representing LA in the DBpedia dataset[10] being linked with co-references in non-English DBpedias[11] or co-references in other datasets such as Freebase.[12]

These principles have enabled the emergence of a global graph of LD on the Web, including cross-domain data such as DBpedia, WordNet RDF,[13] or the data.gov.uk initiative, as well as domain-specific expert vocabularies, for instance, of data about cultural heritage (e.g., the Europeana dataset[14]). This has also led to the creation of an embryonic "Web of Educational Data" (see d'Aquin et al., 2013; Taibi, Fetahu, & Dietze, 2013; and Dietze et al., 2013, for an overview) including data from institutions such as the Open University (UK)[15] or the National Research Council (Italy),[16] as well as publicly available educational resources, such as the mEducator – Linked Educational Resources (Dietze, Taibi, Yu, & Dovrolis, 2015). Initiatives such as LinkedEducation.org,[17] LinkedUniversities.org,[18] and LinkedUp[19] have provided first efforts to bring together people and works in this area. In this context, the LinkedUp Catalog[20] is an unprecedented collection of publicly available LD relevant to educational scenarios, containing data about dedicated open educational resources (OER), such as Open Courseware (OCW) or mEducator datasets, data about bibliographic resources, or metadata about other knowledge resources.

While data about learning activities is not frequently available and data sharing even less so, LD has been adopted to facilitate representation of social and activity or attention data (Dietze, Drachsler, & Giordano, 2014); Ben Ellefi, Bellahsene, Dietze, and Todorov (2016) provide a thorough overview. In the field of LA, LD principles can substantially improve the disambiguation, interpretation, and understanding of data (as documented by d'Aquin & Jay, 2013; d'Aquin et al., 2014). Reference knowledge graphs, domain-specific or cross-domain, can significantly improve the interpretation and analytical processes of captured learning analytics data by disambiguating and enriching data, for instance, about subjects or competencies. This can improve the performance of learning analytics methods and tools within specific scenarios (d'Aquin & Jay, 2013).

Certain limitations are apparent, however, when dealing with reasoning-based approaches such as Semantic Web technologies. Given the computational demands of interpreting and reasoning on knowledge representations, LD-based approaches are known to be less scalable than traditional RDBMS-based[21] methods. However, given the maturity of existing RDF storage and reasoning engines, this applies specifically to very large-scale datasets, which are less frequent in LA and EDM settings. Other issues include the lack of links, the misuse of schema terms, or the lack of semantic and syntactic quality of exposed data. However, these issues are by no means exclusive or specific to LD-based datasets but prevail across data management

[7] http://lak.linkededucation.org
[8] http://www.w3.org/TR/2014/NOTE-rdf11-primer-20140624/
[9] http://www.w3.org/TR/rdf-sparql-query/
[10] http://dbpedia.org/page/Learning_analytics
[11] http://fr.dbpedia.org/resource/Analyse_de_l'apprentissage
[12] http://rdf.freebase.com/ns/m.0crfzwn
[13] http://www.w3.org/TR/2006/WD-wordnet-rdf-20060619/
[14] http://ckan.net/package/europeana-lod
[15] http://data.open.ac.uk
[16] http://data.cnr.it
[17] http://linkededucation.org
[18] http://linkeduniversities.org
[19] http://linkedup-project.eu
[20] http://data.linkededucation.org/linkedup/catalog/
[21] Relational database management system.

Table 29.1. Edited Excerpt of Discourse Data Coded in ENA Format

Publication Venue	# Papers	Type	Named Graph URI
Proceedings of the ACM International Conference on Learning Analytics and Knowledge (LAK) (2011–2014)	166	ACM	http://lak.linkededucation.org/acm http://lak.linkededucation.org/acm/body
Proceedings of the International Conference on Educational Data Mining (EDM) (2008–2014)	463	Open Access	http://lak.linkededucation.org/openaccess http://lak.linkededucation.org/openaccess/body
Special 2012 issue on "Learning and Knowledge Analytics" edited by George Siemens & Dragan Gašević: Educational Technology & Society, 15(3), 1–163.	10	Open Access	
Journal of Educational Data Mining (2009–2014)	29	Open Access	
Journal of Learning Analytics (2014)	16	Open Access	
Proceedings of the LAK data Challenge (2013–2014)	13	Open Access	

technologies of all kinds.

Hence, sharing LA data according to LD principles has the potential to boost the adoption and improvement of LA tools and methods significantly by enabling their evaluation across a range of real-world datasets and scenarios.

THE LAK DATASET: A LINKED DATA CORPUS FOR THE LEARNING ANALYTICS COMMUNITY

In order to provide a best-practice example of adopting LD principles for sharing LA data, we introduce the LAK dataset, a joint effort of an international consortium consisting of the Society for Learning Analytics Research (SoLAR), ACM, the LinkedUp project, and the Educational Technology Institute of the National Research Council of Italy (CNR-ITD). The LAK dataset constitutes a near complete corpus of collected research works in the areas of LA and EDM since 2011, where LD principles have been applied to expose both metadata and full texts of articles (Dietze, Taibi, & d'Aquin, 2017). As such, the corpus enables unprecedented research on the scope and evolution of the LA community. Here, Table 29.1 reports an overview of the publications included in the LAK dataset. Given the variety of sources, the data is split into four subgraphs (last column of Table 29.1 where different license models apply).[22]

To ensure wide interoperability of the data, we have adapted LD best practices[23] and investigated widely used vocabularies for the representation of scientific publications. The scope of our data model is not covered by a single vocabulary alone. For this reason, we opted for using established vocabularies such as BIBO, FOAF,[24] SWRC, and Schema.org for all represented terms and included mappings between the chosen vocabularies as well as other overlapping ones. [25] The choice of vocabulary terms was influenced by the Web-wide adoption and maturity of the used schemas and their overlap with our data model. Table 29.2 reports the concepts represented in the LAK dataset and their population while Table 29.3 summarizes the most frequently populated properties.

Table 29.2. Entity Population in the LAK Dataset

Concept	Type	#
Reference	schema:CreativeWork	7885
Author	swrc:Person	1214
Conference Paper	swrc:InProceedings	697
Organization	swrc:Organization	365
Journal Paper	swrc:Article	45
Conference Proceedings	swrc:Proceedings	15
Journal Issue	bibo:Issue	9
Journal	bibo:Journal	2

Exploiting inherent features of LD, the LAK dataset is enriched with entity links to other datasets, for instance to provide links to author and publication venue co-references and complementary information. In particular, links with the Semantic Web Dog Food (SWDF)[26] dataset and DBLP provide additional information about authors and venues in the LAK dataset,

[22] Data from graphs http://lak.linkededucation.org/openaccess/* are available under CC-BY licence. For data in graphshttp://lak.linkededucation.org/acm/*, we have negotiated a formal agreement with ACM to publish, share, and enable reuse of the data for research purposes. https://creativecommons.org/licenses/by/2.0/

[23] http://www.w3.org/TR/ld-bp/#VOCABULARIES

[24] http://xmlns.com/foaf/spec/

[25] The currently implemented schema is available at http://lak.linkededucation.org/schema/lak.rdf While this URL always refers to the latest version of the schema, current and previous versions are also accessible, for instance, via http://lak.linkededucation.org/schema/lak-v0.2.rdf

[26] http://data.semanticweb.org/

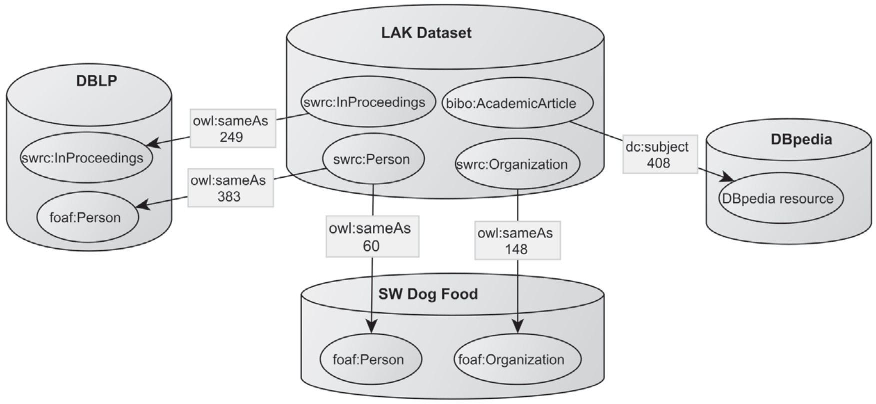


Figure 29.1. Interlinking the LAK dataset.

such as their wider scientific activity and impact. This is useful, for instance, to complement the highly focused nature of the LAK dataset, which by definition has a narrow scope (LA and EDM) and would otherwise limit research to activities within that very community. On the other hand, such established links complement existing corpora with data contained in the LAK dataset by 1) enriching the limited metadata with additional properties and 2) containing additional publications not reflected in DBLP or the Semantic Web Dog Food, creating a more comprehensive knowledge graph of Computer Science literature as a whole.

Table 29.3. Entity Population in the LAK Dataset

Domain	Property	Range	#
schema:Article	schema:citation	schema:CreativeWork	10828
swrc:InProceed-ings	dc:subject	literal	3392
foaf:Agent	foaf:made	swrc:InProceedings	2199
foaf:Person	rdfs:label	literal	1583
foaf:Agent	foaf:sha1sum	literal	1341
swrc:Person	swrc:affiliation	swrc:Organization	1293
foaf:Person	foaf:based_near	geo:SpatialThing	1243
schema:Article	schema:article-Body	literal	698
bibo:Article	bibo:abstract	literal	697
bibo:Issue	bibo:hasPart	bibo:Article	45
swrc:Proceedings	swc:relatedToEvent	swc:ConferenceEvent	14
bibo:Journal	bibo:hasPart	bibo:Issue	9

Additional outlinks were created to DBpedia as reference vocabulary. To allow for a more structured retrieval and clustering of publications according to their topic-wise similarity, we have linked *keywords*, provided by authors, to their corresponding entities in DBpedia, thereby using DBpedia as reference vocabulary for paper topic annotations. Figure 29.1 depicts the links of resolved or enriched LAK entities.

Given the nature of LD, establishing such links has been merely a matter of looking up LAK dataset entities and adding *owl:sameAs* statements, which refer to the IRIs of co-referring entities in DBLP, SW Dog Food, and DBpedia. Hence, this process is enabled by fundamental principles of LD, such as using URIs to identify things and using SPARQL queries to demonstrate the key motivation: the creation of a global data graph rather than isolated datasets.

LINKED DATA-ENABLED INSIGHTS INTO THE LAK CORPUS: SCOPE AND TRENDS OF LEARNING ANALYTICS RESEARCH

To illustrate the exploitation of LD principles implemented in the LAK dataset, we introduce some simple analysis enabled through the inherent links within the dataset, as described above. While a wide range of additional investigations can be found in the applications and publications of the LAK Data Challenge,[27] here we focus on a set of very simple investigations and research questions. These can be answered merely by combining SPARQL queries on the LAK dataset and interlinked datasets, and by exploiting the links between co-references described in the earlier section. These analyses are primarily aimed at demonstrating the ease of answering complex research questions by combining data from different sources. In particular, we investigate questions related to the following:

1. The research background and focus of researchers in the LA community, in order to shape a picture of the constituting disciplines and areas of this comparably new research area: This investigation

[27] See the application and publication sections at http://lak.linkededucation.org

Table 29.4. Top 20 Conferences and Journals in which LAK 2011 Conference Authors Previously Published

Conference or Journal	DBLP resource	%
Intelligent Tutor Systems Conference	http://dblp.l3s.de/d2r/resource/conferences/its	24.49
Educational Data Mining Conference	http://dblp.l3s.de/d2r/resource/conferences/edm	12.05
Artificial Intelligence in Education Conference	http://dblp.l3s.de/d2r/resource/conferences/aied	11.15
European Conference on Technology Enhanced Learning	http://dblp.l3s.de/d2r/resource/conferences/ectel	7.31
International Conference on Advanced Learning Technologies	http://dblp.l3s.de/d2r/resource/conferences/icalt	5.51
AAAI Conference on Artificial Intelligence	http://dblp.l3s.de/d2r/resource/conferences/aaai	4.36
UM conference	http://dblp.l3s.de/d2r/resource/conferences/um	4.36
IEEE International Conference on Data Mining	http://dblp.l3s.de/d2r/resource/conferences/icdm	3.72
Conference on Knowledge Discovery and Data Mining	http://dblp.l3s.de/d2r/resource/conferences/kdd	3.59
International Journal of Artificial Intelligence in Education	http://dblp.l3s.de/d2r/resource/journals/aiedu	2.56
ETS journals	http://dblp.l3s.de/d2r/resource/journals/ets	2.31
Conference on Computer Supported Collaborative Learning	http://dblp.l3s.de/d2r/resource/conferences/cscl	2.31
International Conference on Machine Learning	http://dblp.l3s.de/d2r/resource/conferences/icml	2.31
Journal of Universal Computer Science	http://dblp.l3s.de/d2r/resource/journals/jucs	2.18
International Conference of the Learning Sciences	http://dblp.l3s.de/d2r/resource/conferences/icls	2.18
ACM CHI Conference	http://dblp.l3s.de/d2r/resource/conferences/chi	2.18
World Wide Web conference	http://dblp.l3s.de/d2r/resource/conferences/www	2.18
AH conference	http://dblp.l3s.de/d2r/resource/conferences/ah	1.92
International Joint Conference on Artificial Intelligence	http://dblp.l3s.de/d2r/resource/conferences/ijcai	1.67
Machine Learning Journal	http://dblp.l3s.de/d2r/resource/journals/ml	1.67

exploits information about LA researchers' publication activity in other areas by using DBLP data.

2. The importance of key topics in the LA field and their evolution over time: This investigation exploits the topic (or category) mapping of LA keywords (or entities) in DBpedia and their relationships.
3. The apparent links between the LA and LD communities that can be derived from the data.

In both cases, our analysis has been conducted by taking into account all the publications of the LAK conferences from 2011 to 2014, available in the LAK dataset, in order to study the evolution of LA research over the years.

Who Makes Up the Learning Analytics Community?Publication Activities of LA Researchers

The development of LA has been influenced by the intersection of numerous academic disciplines such as machine learning, artificial intelligence, education technology, and pedagogy (Dawson, Gašević, Siemens, & Joksimović, 2014). For this reason, since its first edition, the LAK conference has drawn the attention of researchers from different scientific fields, each contributing their definitions, terminologies, and research methods, and thereby shaping the definition of what LA is. The core data of the LAK dataset, being limited to LA-related publication activities exclusively, does not enable any analysis into the origin and research background of contributing researchers. The LD nature of the corpus, however, provides meaningful connections that can be exploited to infer such new knowledge. In fact, by linking the resources representing the authors in the LAK dataset with the authors in the DBLP dataset, it is feasible with a few SPARQL queries to extract further information about the fields of interest of the authors.[28]

For all LAK authors in each year from 2011 to 2014, we analyzed the number of publications in previous conferences and journals by first 1) obtaining all authors of a respective year in the LAK dataset and 2) retrieving their previous publication venues (journals,

[28] The 36% of the 1214 authors in the LAK dataset is linked with the correspondent resource in the DBLP (86%) and SWDF datasets (14%).

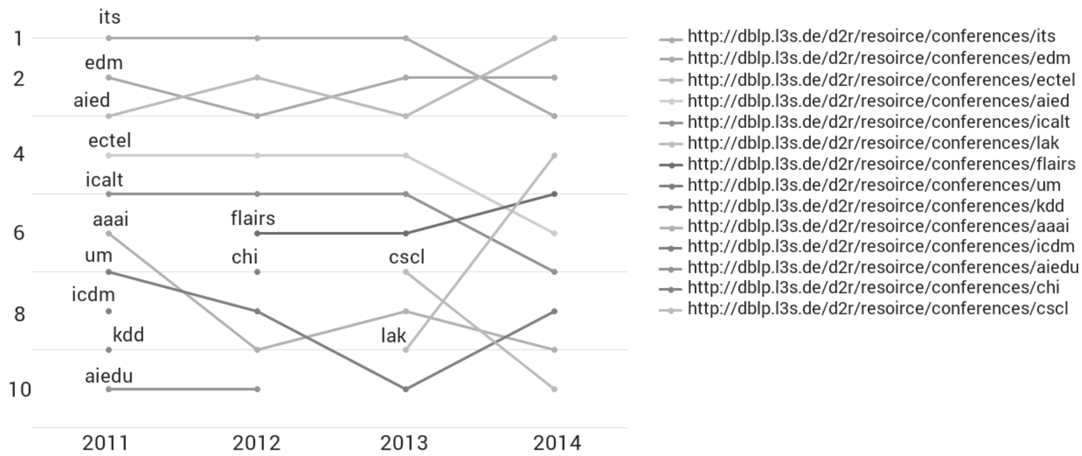


Figure 29.2. Interlinking the LAK dataset.

conferences) from DBLP. The top 20 conferences and journals for 2011 are reported in Table 29.4. This table highlights that, in its first edition, the LAK conference has mainly involved authors with previous publications related to the *Intelligent Tutor Systems*, *Educational Data Mining*, *Artificial Intelligence*, and *Technology Enhanced Learning* conferences. From a technical point of view, interlinks between the LAK and DBLP datasets were created as follows: LAK authors are linked with their co-references in the DBLP dataset through the *owl:sameAs* property The DBLP authors in turn are connected with their publications in previous conferences and journals respectively through the *swrc:series* and the *swrc:journal* properties. The execution of a federated query involving the two datasets allows us to deduce information about the number of publications of LAK authors in previous conferences and journals.

In Figure 29.2, we report the rank of the top 10 conferences and journals in which LAK authors have published from 2011 to 2014. The top three positions are clearly occupied by the ITS (Intelligent Tutor Systems), EDM (Educational Data Mining), and AIED (Artificial Intelligence in Education) conferences/journals. Starting in 2013, the LAK conference appears in the top 10, growing in importance in 2014, indicating the constitution of a significant community in its own right. That same year saw an increasing number of papers published in the FLAIRS (Florida Artificial Intelligence Research Society) conference proceedings. Publications from EC-TEL (European Conference on Technology Enhanced Learning) and ICALT (International Conference on Advanced Learning Technologies) conferences also appear at the top of the list, with a slight inflection in the last year.

Which Topics Make Up the LA Field? How Does Topic Distribution Change Over Time?

In contrast to the previous investigations, interlinking LAK conference publications with relevant DBpedia

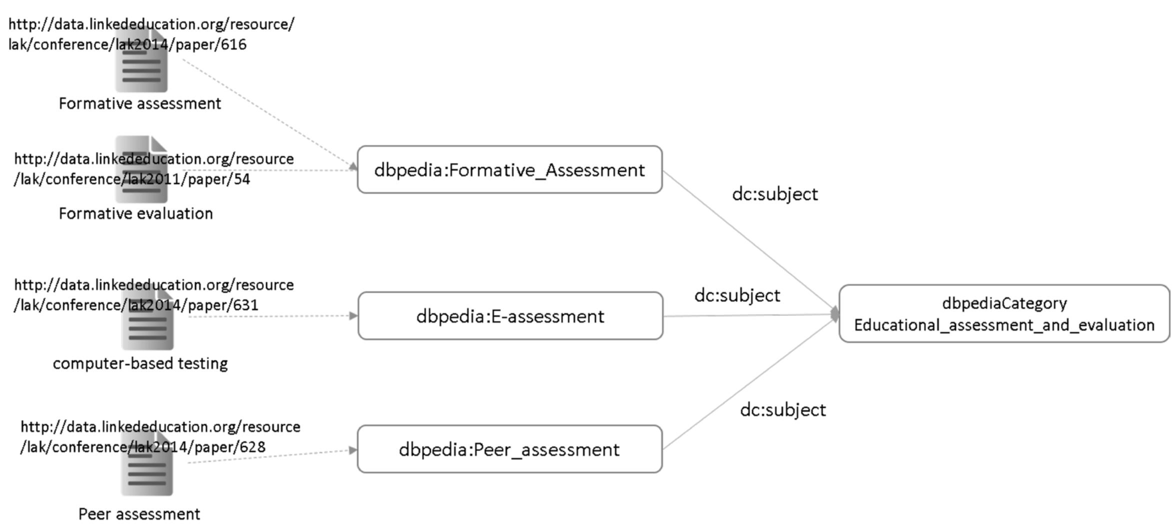


Figure 29.3. Interlinking publications and DBPedia.

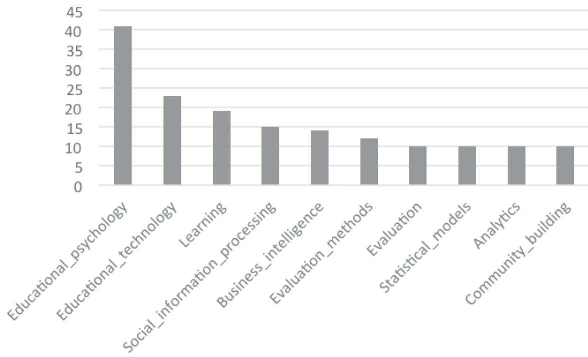


Figure 29.4. Top-10 DBPedia categories for the LAK dataset (publications from 2011-2014).

entities allows us to investigate the semantics of topics covered by analyzing the DBpedia knowledge graph and the inherent links of entities and categories. This, for instance, enables us to identify the overlap of LAK papers with other disciplines, such as *Computer Science*, *Statistics*, *Technology Enhanced Learning*, or *Data Analysis*.

As described in the previous section, links between the LAK dataset and DBpedia entities were established by disambiguating terms (keywords) through state-of-the-art NER (Name Entity Recognition) methods (DBpedia Spotlight). This allowed us to link keywords – for instance, "computer-based testing" and "formative

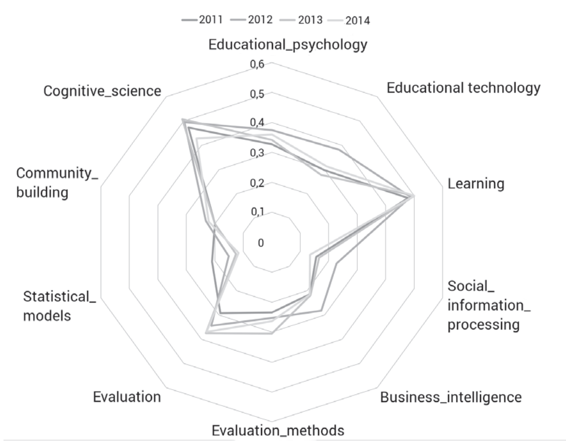


Figure 29.5. Evolution of the top-10 categories over time.

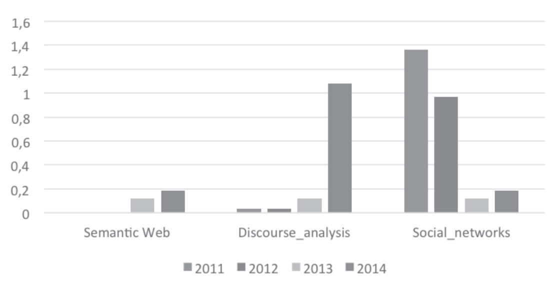


Figure 29.6. Evolution of selected categories, 2011–2014.

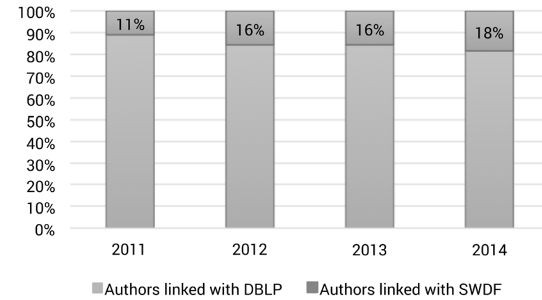


Figure 29.7. Percentage of LAK authors represented within DBLP and Semantic Web Dog Food.

evaluation" – respectively to the corresponding DBpedia entities, *http://dbpedia.org/resource/E-assessment* and *http://dbpedia.org/resource/Formative_assessment*[29] (see Figure 29.3). Each DBpedia entity, in turn, is connected through the dc:subject property, to its corresponding DBpedia categories; for instance, the category Educational_assessment_and_evaluation is the dc:subject of DBpedia resources: Formative_assessment, E-assessment, Peer_assessment, and Educational_evaluation, just to name a few. In this way, papers can be clustered according to their structural similarity within the DBpedia graph. The list of top 10 DBpedia categories with the highest frequency value in the LAK dataset is shown in Figure 29.4.[30]

Starting from this set of top-10 most frequent categories over 2011 to 2014, we evaluated the distances between all the DBpedia categories extracted for each conference year and the categories included in this set of "base categories." The SKOS[31] properties used by the DBpedia category graph to represent relationships between categories were exploited to compute this distance. For example, the distance between *E-Learning* and *Educational_technology* is 2, since *Educational_technology* is *skos:broader* of *Distance_education* and, in turn, *Distance_education* is *skos:broader* of *E-learning*.

The relation between DBpedia categories and LAK conference papers also makes it easier to trace the trend of topics covered by LAK publications over the years. The radar chart in Figure 29.5 provides an overview of the calculated average distance between *all categories extracted for each conference year and each category included in the "base category" set*. From the analysis of the figure, the following considerations arise:

- *Educational_technology* played a key role in 2012 but in other years more specialized categories

[29] See the following papers: http://data.linkededucation.org/resource/lak/conference/lak2011/paper/54 and http://data.linkededucation.org/resource/lak/conference/lak2014/paper/616

[30] References to DBpedia categories are in the form: http://dbpedia.org/page/Category:Educational_psychology, http://dbpedia.org/page/Category:Educational_technology and so on.

[31] http://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html

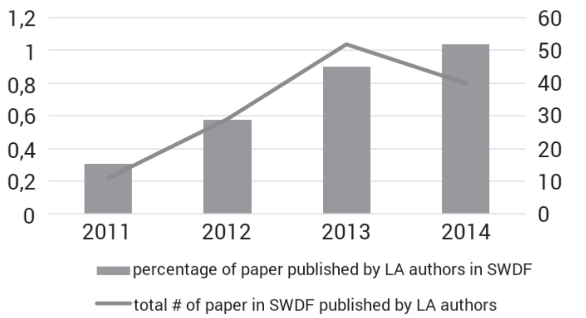


Figure 29.8. Trend of previous publications of LAK authors in SW conferences.

gained importance

- The fairly broad categories of *Learning* and *Evaluation* have had the greatest relevance in all years
- The relevance of *Evaluation* has an increasing trend over the years, with a sensitive increment from 2011 to 2012.
- *Statistical_models* peaked in 2011 and decreased in subsequent years.

To better understand trends for selected categories, Figure 29.6 reports the normalized frequency, calculated for three arbitrarily selected categories, as the actual number of occurrences of a particular category minus the mean of all frequencies divided by the standard deviation. The *Semantic_Web* category appeared in LAK publications in 2013 and a slight increment can be observed between 2013 and 2014. The analysis of the trend for the *Discourse_Analysis* category reveals a positive increment over the years with a remarkable increment registered in the last year. On the contrary, we observe a negative trend for the *Social_networks* category; in fact, the relevance of this category decreased substantially from 2011 to 2013, with a slightly increment in 2014.

Is There a Link Between the LD and LA Communities?

As indicated above, the analysis of authors contributing to the LAK community and the topic coverage of LAK publications provides clues about the influence of Semantic Web on researchers in the LAK community, a question of relevance to the scope of this article. Figure 29.7 shows the percentage of authors linked with either DBLP or the Semantic Web Dog Food dataset, showing a positive trend related to the increment of authors from the Semantic Web community. This can be attributed either to SW researchers publishing more strongly in the LA community or that LA researchers began publishing in SW-related venues.

To investigate this further, the links between the authors of the LAK dataset and the Semantic Web Dog Food have been exploited to determine the number of Semantic Web-related publications by LA authors. These have been measured by the *number of publications in the SWDF dataset by LA authors*. As we already know from Figure 29.2, SW conferences are not in the top 10 list of previous publications for LAK authors, but the percentage of papers published by LA authors in SW conferences shows a positive trend over the years, even if the total number reduced in 2014, as reported in Figure 29.8.

While some of these insights are hardly surprising, the ease with which they could be generated is worth highlighting: in all cases, data was fetched with a few SPARQL queries, where the previously established links between co-references across different datasets (LAK, DBpedia, DBLP, SWDF) allows the correlation of data from these different sources to answer more complex questions.

CONCLUSIONS AND LESSONS LEARNED

Applying LD principles when dealing with LA data, or any kind of data, has benefits specifically for understanding and interpreting data. As a key component of LD principles, one of the enabling building blocks is the use of global URIs for identifying entities and schema terms across the Web, which provides the foundations for cross-dataset linkage and querying, essentially creating a global knowledge graph.

In order to demonstrate the opportunities arising from adopting LD principles in LA and present some insights into the state and evolution of the LA community and discipline, we have introduced the LAK dataset, together with a set of example questions and insights. These include investigations into the composition of the LA community as well as the significant topics and trends that can be derived from the LAK dataset when considering other LD sources, such as DBLP or DBpedia, as background knowledge.

While these insights were not meant to provide a thorough investigation of the state of the LA field, they provide a glimpse into the opportunities arising from following LD principles and exploiting external data sources for interpreting data and investigating more complex research questions, which would not be feasible by looking at isolated data sources.

In this regard, a number of best practices emerge when sharing and reusing data on the Web, concerning 1) the data publishing side and 2) the data analysis side. Regarding the former, previous work (Dietze, Taibi, & d'Aquin, 2017) describes the practices and design choices applied when building and publishing the LAK dataset. Here, next to the general LD principles, we paid particular attention to designing a schema from established and well-used vocabulary terms. We considered a range of criteria, including the wide adoption of the used vocabulary terms, their coverage and match with the data model of the LAK dataset, as well as their inherent compatibility. We applied similar criteria when choosing linking candidates, such as DBLP or DBpedia, to enable more meaningful analysis of the LA community and its scientific output. While finding candidate datasets for the linking task is an inherently difficult problem, automated approaches (Ben Ellefi et al., 2016) can be applied to aid dataset providers.

While our initial analysis of the LAK dataset only provided a limited perspective on certain aspects of the LAK community and its evolution, it illustrates the ease with which particular research questions can be investigated using a well-defined and interlinked dataset, as opposed to a traditional database. More thorough studies of the LA community have been carried out as part of the LAK Data Challenge, in which researchers have been invited to develop applications aimed at providing innovative exploration of the data contained in the LAK dataset.

REFERENCES

d'Aquin, M., Adamou, A., & Dietze, S., (2013). Assessing the educational linked data landscape. *Proceedings of the 5th Annual ACM Web Science Conference* (WebSci '13), 2–4 May 2013, Paris, France (pp. 43–46). New York: ACM.

d'Aquin, M., & Jay, N. (2013). Interpreting data mining results with linked data for learning analytics: Motivation, case study and direction. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 155–164). New York: ACM.

d'Aquin, M., Dietze, S., Herder, E., Drachsler, H., & Taibi, D. (2014). Using linked data in learning analytics. *eLearning Papers*, 36. http://hdl.handle.net/1820/5814

Ben Ellefi, M., Bellahsene, Z., Dietze, S., & Todorov, K. (2016). Intension-based dataset recommendation for data linking. In H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S. P. Ponzetto, & C. Lange (Eds.), *The Semantic Web: Latest Advances and New Domains* (pp. 36–51; Lecture Notes in Computer Science, Vol. 9678). Springer.

Bizer, C., Heath, T., & Bernes-Lee, T. (2009). Linked data: The story so far. https://wtlab.um.ac.ir/images/e-library/linked_data/Linked%20Data%20-%20The%20Story%20So%20Far.pdf

Dawson, S., Gašević, D., Siemens, G., & Joksimović, S. (2014). Current state and future trends: A citation network analysis of the learning analytics field. *Proceedings of the 4th International Conference on Learning Analytics & Knowledge*(LAK '14), 24–28 March 2014, Indianapolis, IN, USA (pp. 231–240). New York: ACM.

Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H., Giordano, D., Marenzi, I., & Pereira Nunes, B. (2013). Interlinking educational resources and the web of data: A survey of challenges and approaches. *Program: Electronic Library and Information Systems*, 47(1), 60–91.

Dietze, S., Taibi, D., Yu, H. Q., & Dovrolis, N. (2015). A linked dataset of medical educational resources. *British Journal of Educational Technology*, 46(5), 1123–1129.

Dietze, S., Taibi, D., & d'Aquin, M. (2017). Facilitating scientometrics in learning analytics and educational data mining: The LAK dataset. *Semantic Web Journal*, 8(3), 395–403.

Dietze, S., Drachsler, H., & Giordano, D. (2014). A survey on linked data and the social web as facilitators for TEL recommender systems. In N. Manouselis, K. Verbert, H. Drachsler, & O. C. Santos (Eds.), *Recommender systems for technology enhanced learning: Research trends and applications* (pp. 47-77). Springer.

Heath, T., & Bizer, C. (2011). *Linked data: Evolving the web into a global data space*. San Rafael, CA: Morgan & Claypool Publishers.

Taibi, D., Fetahu, B., & Dietze, S. (2013). Towards integration of web data into a coherent educational data graph. *Proceedings of the 22nd International Conference on World Wide Web* (WWW '13), 13–17 May 2013, Rio de Janeiro, Brazil (pp. 419–424). New York: ACM. doi:10.1145/2487788.2487956

Chapter 30: Linked Data for Learning Analytics: Potentials and Challenges

Amal Zouaq[1], Jelena Jovanović[2], Srećko Joksimović[3], Dragan Gašević[2,4]

[1] School of Electrical Engineering and Computer Science, University of Ottawa, Canada
[2] Department of Software Engineering, University of Belgrade, Serbia
[3] Moray House School of Education, University of Edinburgh, United Kingdom
[4] School of Informatics, University of Edinburgh, United Kingdom

DOI: 10.18608/hla17.030

ABSTRACT

Learning analytics (LA) is witnessing an explosion of data generation due to the multiplicity and diversity of learning environments, the emergence of scalable learning models such as massive open online courses (MOOCs), and the integration of social media platforms in the learning process. This diversity poses multiple challenges related to the interoperability of learning platforms, the integration of heterogeneous data from multiple knowledge sources, and the content analysis of learning resources and learning traces. This chapter discusses the use of linked data (LD) as a potential framework for data integration and analysis. It provides a literature review of LD initiatives in LA and educational data mining (EDM) and discusses some of the potentials and challenges related to the exploitation of LD in these fields.

Keywords: Linked data (LD), data integration, content analysis, educational data mining (EDM)

The emergence of massive open online courses (MOOCs) and the open data initiative have led to a change in the way educational opportunities are offered by shifting from a university-centric model to a multi-platform and multi-resource model. In fact, today's learning environments include not only diverse online learning platforms, but also social media applications (e.g., SlideShare, YouTube, Facebook, Twitter, or LinkedIn) where learners connect, communicate, and exchange data and resources. Henceforth, learning is now occurring in various forms and settings, both at the formal (university courses) and informal (social media, MOOC) levels. This has led to a dispersion of learner data across various platforms and tools, and brought a need for efficient means of connecting learner data across various environments for a comprehensive insight into the learning process. One salient example of the need for data exchange across platforms is the connectivist MOOC (cMOOC). In cMOOCs, learning, by definition, does not take place in a single platform, but relies on a range of dedicated online learning applications as well as social media and networking applications for sharing information and resources among learners (Siemens, 2005). These developments led to new requirements and imposed new challenges for both data collection and use.

From the perspective of data collection, the emergence of cloud services and the rapid development of scalable web architectures allow for pulling and mashing data from various online applications. This is supported by the development of large-scale interfaces (APIs) by major Web stakeholders such as Facebook, LinkedIn, or Twitter, and by MOOC providers such as Coursera and Udacity. From the perspective of data use, the plethora of resources and interactions occurring in educational platforms requires analytic capabilities, including the ability to handle different types of data. Various kinds of data are generated, some of which capture learners' interactions in learning and social media platforms (learners' logs/traces), whereas others take the form of unstructured content, ranging from course content and learners' blogs to discussion forum posts. This multitude of kinds and sources of data provides fertile ground for the field of learning analytics and its overall objectives to better understand learners and the learning process, provide timely, informative, and adaptive feedback, and foster lifelong learning (Gaešvić, Dawson, & Siemens, 2015).

Challenges associated with the collection, integration, and use of data originating from heterogeneous sources are often dealt with, in the educational community, by developing a standardized data model that allows for integration and leveraging of heterogeneous data (Dietze et al., 2013). This chapter focuses on linked data (LD) as one potential approach to the development and use of such a data model in both formal and informal online learning settings. In particular, the use of LD principles (Bizer, Heath, & Berners-Lee, 2009) allows for establishing a globally usable network of information across learning environments (d'Aquin, Adamou, & Dietze, 2013), leading to a global educational graph. Similar graphs could be created at the individual level, for each particular learner, connecting all the data and resources associated with their learning activities. The educational potentials and benefits of such graphs have already been examined and discussed. For instance, Heath and Bizer (2011) propose an *educational graph* across UK universities, comprising knowledge extracted from the content of learning resources. Given the development and use of knowledge graphs by an increasing number of major companies such as Google, Microsoft, and Facebook, the potential and possibilities opened up by such graphs for learning should be examined (Zablith, 2015).

This chapter describes the current state of the art of LD usage in education, focusing primarily on existing and potential applications in the learning analytics (LA)/educational data mining (EDM) field. After a brief introduction to LD principles in the next section, the chapter analyzes the potential of LD along two particular dimensions: 1) the data integration dimension and 2) the data analysis and interpretation dimension. Finally, we discuss some potentials and challenges associated with the use of LD in LA/EDM.

LINKED DATA IN EDUCATION

Linked data has the potential to become a de facto standard for sharing resources on the Web (Kessler, d'Aquin, & Dietze, 2013). It uses URIs to uniquely identify entities, and the RDF data model[1] to describe entities and connect them via links with explicitly defined semantics. In particular, LD relies on four principles:

1. Use URIs as names for things; for instance, historical novel "Paris" is uniquely identified by its ISBN (a kind of URI): 0385535309
2. Provide the ability to look up names through HTTP URIs; while an ISBN does uniquely identify a book, it cannot be used to provide direct access to it on the Web, so HTTP URIs should be used instead; the book from our example could be looked up via the following HTTP URI: <http://www.worldcat.org/oclc/827951628>
3. Upon URI look up, return useful information using the standards RDF and SPARQL[2]; for instance, we can state, in a machine-processable manner, that the resource identified by the <http://www.worldcat.org/oclc/827951628> URI is of the *type* book and belongs to the genre of historical fiction: *<http://www.worldcat.org/oclc/827951628> rdf:type schema:Book* ; *schema:genre "Historical fiction"*.
4. Include links to other entities uniquely identified by their URIs; for instance, we can connect the book from our example with its author: *<http://www.worldcat.org/oclc/827951628> schema:author <http://viaf.org/viaf/*34666*>* where the latter URI uniquely identifies the writer Edward Rutherfurd.

Thanks to the simplicity of these principles, LD represents an elegant framework for modelling and querying data at a global scale. It is usable in various applications and domains, and can constitute a response to the interoperability and data management challenges that have long faced the educational community (Dietze et al., 2013).

Billions of data items have been published on the web as linked data, forming a global open data space – the linked open data cloud (LOD)[3] – that includes open data from various domains such as government data, scientific knowledge, and data about a variety of online communities. Huge cross-domain knowledge bases have also emerged on the LOD such as DBpedia[4], Yago[5], and Wikidata[6]. As such, LD has the potential to enable a global shift in how data is accessed and utilized, offering access to data from various sources, through various kinds of data access points, including Web services and Web APIs, and allowing for seamless creation of dynamic data mashups (Bizer et al., 2009). In fact, one salient feature of LD is that it establishes semantic-rich connections between items from different data sources, and thus opens up data silos (e.g., traditional databases) for more seamless data integration and reuse.

Despite all these potential benefits, the LD formalism and technologies have had a slow adoption in the area of technology-enhanced learning; initiatives that employ LD technologies have only emerged recently (Dietze et al., 2013). We can identify several application scenarios

[1] Resource Description Framework, http://www.w3.org/RDF/

[2] https://www.w3.org/TR/sparql11-query/
[3] http://lod-cloud.net/
[4] http://lod-cloud.net/
[5] http://bit.ly/yago-naga
[6] http://bit.ly/wikidata-main

in the LA/EDM field that would benefit from the LOD, including 1) resource discovery (e.g., faceted search) and content enrichment (e.g., augmenting content with data from LOD datasets) (Maturana, Alvarado, López-Sola, Ibáñez, & Elósegui, 2013); 2) content analysis based on semantic annotation (Joksimovi et al., 2015); 3) resource and service integration (Dietze et al., 2012); 4) personalization (Dietze, Drachsler, & Giordano, 2014); and 5) interpretation of EDM results (d'Aquin et al., 2013).

DATA INTEGRATION USING LINKED DATA

One of the most salient benefits of LD lies in its data integration potential. This is particularly relevant for the LA/EDM field since it requires the collection and management of learner and content data from a variety of sources (applications and services) used in informal and life-long learning (Santos et al., 2015). In particular, to build a comprehensive learner model, one needs to integrate learner data recorded in different learning platforms/tools the learner has interacted with (Desmarais & Baker, 2012). Therefore, the challenges associated with handling multiple data formats and the overall lack of data interoperability, are becoming a key issue (Chatti, Dyckhoff, Schroeder, & Thüs, 2012; Duval, 2011). More generally, the ease of data transfer, pre-processing, use, combination and analysis without loss of meaning across learning platforms are becoming important factors for the efficiency of LA/EDM (Cooper, 2013).

Several domains have been successful in exploiting LD for data integration issues such as the biomedical domain (Belleau, Nolin, Tourigny, Rigault, & Morissette, 2008), pharmacology (Groth et al., 2014), and environmental sciences (Lausch, Schmidt, & Tischendorf, 2015). All of this suggests that LD technologies could provide the solid data integration layer that LA/EDM necessitates.

Previous Initiatives in Data Integration in the Educational Community

The technology enhanced learning research community has long recognized the importance of data integration, which eventually resulted in multiple standardization efforts. Cooper (2013) provides a valuable overview of various standards related to learning. Mainly, these standards relate to the representation of data about learners and their activities, as well as learning content and services.

At the learner level, standards focus on facts about individuals and their history, their connections and interactions with other persons, and interactions with resources offered by learning environments (person and learning activities dimensions). Various specifications exist to model learners (e.g., FOAF[7]), and learner activities and interactions (e.g., Contextualized Attention Metadata [Schmitz, Wolpers, Kirschenmann, & Niemann, 2011], Activity Streams[8], or ADL xAPI[9]).

At the content level, previous initiatives such as IEEE Learning Object Metadata (LOM)[10] and ADL SCORM[11] attempted to create vocabularies and standards that would unify the description of online educational resources or the specification of computer-based assessment (e.g., IMS QTI[12]). Other efforts targeted the mapping between various data models, such as the work of Niemann, Wolpers, Stoitsis, Chinis, and Manouselis (2013) who aimed at aggregating sets of social and interaction data. Finally, several interfaces were proposed to provide guidelines for the implementation of services compliant with these standards (Dietze et al., 2013).

Based on different viewpoints, these efforts led to multiple competing projects and thus created sub-communities with various technologies, languages, and models, and very little interoperability among them. The LD philosophy provides a solution to these interoperability issues by allowing a multiplicity of models on the Web, bridging these models using Web-accessible semantic links. Thus semantically similar models that are differently represented can still be aligned using typed links that establish meaningful connections between concepts originating from different models; for instance, equality connections (*owl:sameAs*), or hierarchical connections (*rdfs:subclassOf or skos:broader*).

Current Data Integration Initiatives Using Linked Data

Integration based on LD requires the availability of Web-accessible LD vocabularies that describe the types of entities in specific subject domains, entities' attributes, and the kinds of connections among the entities. It also depends on the availability of services that allow for exploiting multiple datasets for a given task, as well as services that expose data as LD. This section introduces some of the available vocabularies in the educational domain, and efforts aimed at exposing educational data as LD. A more comprehensive overview of education-related vocabularies can be found in Dietze et al. (2014). The section also gives examples of services exploiting the integration of multiple LD datasets.

An increasing number of educational institutions have been exposing their data following LD principles, such as the Open University in the UK or the University

[7] http://www.foaf-project.org/
[8] http://activitystrea.ms/
[9] http://www.adlnet.gov/tla/experience-api
[10] http://ieeeltsc.org/wg12LOM/
[11] http://www.adlnet.org/
[12] http://www.imsglobal.org/question/

of Münster[13] in Germany. One prominent effort in exposing educational data as LD was the LinkedUp project[14], which resulted in a catalog of datasets related to education and encouraged the development of competitions such as the LAK Data Challenge[15], whose aim was to expose LA/EDM publications as LD and promote their analysis by researchers. While these initiatives represent a step in the adoption of LD by the educational community, their impact remains limited. For example, the data representation and use in MOOC platforms — one of the most striking developments in today's technology-enhanced learning — has not been based on LD principles or technologies to date. Still, few recent initiatives (Kagemann & Bansal, 2015; Piedra, Chicaiza, López, & Tovar, 2014) showed some interest in describing and comparing MOOCs using an LD approach. For example, MOOCLink (Kagemann & Bansal, 2015) aggregates open courseware as LD and exploits these data to retrieve courses around particular subjects and compare details of the courses' syllabi. Recently, there has also been an initiative that relies on schema.org[16] to create a vocabulary for course description[17] with the purpose of facilitating the discovery of any type of educational course. Schema.org is a structured data markup schema (or vocabulary) supported by major Web search engines. This schema is then used to annotate Web pages and facilitate the discovery of relevant information. Given its adoption by major players on the Web, this is a welcome initiative that might have some long-term impact in the educational community. Similarly, some authors worked on providing an RDF representation (binding) of educational standards. For example, an RDF binding of the Contextualised Attention Metadata (CAM) (Muñoz-Merino et al., 2010) and an RDF binding of the Atom Activity Streams[18] were developed. This enabled data integration and interoperability both at syntax and semantic levels.

Finally, with the current shift towards RESTful (representational state transfer) services on the cloud, education-related services based on LD have started to emerge. At a conceptual level, we can identify two main types of services based on LD currently being investigated in research: 1) services for course interlinking within a single institution and across institutions, and 2) services for integrating learners' log data based on a common model.

For example, Dietze et al. (2012) proposed an LD-based framework to integrate existing educational repositories at the service and data levels. Zablith (2015) suggested the use of LD as a conceptual layer around higher education programs to interlink courses in a granular and reusable manner. Another work links ESCO[19]-based skills to MOOC course descriptions to create enriched CVs (Zotou, Papantoniou, Kremer, Peristeras, & Tambouris, 2014). Interestingly, the authors are able to identify similar skills taught in the Coursera and Udacity MOOC platforms, thus providing implicit links between courses of two different MOOC platforms. One can envisage exciting opportunities for life-long learning based on a cross-platform MOOC course recommendation service.

Another indicator of the growing importance of LD in the realm of education in general, and LA/EDM in particular, is the adoption of LD concepts and technologies into xAPI specifications[20]. With xAPI, developers can create a learning experience tracking service through a predefined interface and a set of storage and retrieval rules. De Nies, Salliau, Verborgh, Mannens, and Van de Walle (2015) propose to expose data models created using the xAPI specification as LD. This proposal provides an interoperable model of learning traces data, and allows for seamless exposing of learners' traces as semantically interoperable LD. Similarly, Softic et al. (2014) report on the use of Semantic Web technologies (RDF, SPARQL) to model learner logs in personal learning environments.

Based on the scalability of the Web as the base infrastructure, and using the interoperability of the W3C standards RDF and SPARQL, we believe that similar initiatives can further contribute to the development of decentralized and adaptable learning services.

DATA ANALYSIS AND INTERPRETATION USING LINKED DATA

Given the rapid growth of unstructured textual content on various online social media and communication channels, as well as the ever-increasing amount of dedicated learning content deployed on MOOCs, there is a need to automate the discovery of items relevant to distance education, such as topics, trends, and opinions, to name a few. In fact, analytics required for the discovery and/or recommendation of relevant items can be improved if the regular input data (e.g., learners' logs) is enriched with background information from LOD datasets (e.g., data about topics associated with the course) (d'Aquin & Jay, 2013). The use of LOD cross-domain knowledge bases such as DBpedia and Yago, alone or in combination with traditional content analysis techniques (e.g., social network analysis, text mining, latent semantic indexing), represent a promising avenue for advancing content analysis

[13] http://lodum.de/
[14] http://linkedup-project.eu/
[15] http://lak.linkededucation.org/
[16] https://schema.org/
[17] https://www.w3.org/community/schema-course-extend/
[18] http://xmlns.notu.be/aair/
[19] European Commission, "ESCO: European Skills, Competencies, Qualifications and Occupations," https://ec.europa.eu/esco
[20] https://github.com/adlnet/xAPI-Spec

and information retrieval in educational settings, as outlined in the following sections.

Content Analysis Using Semantic Annotation

One important development in the LD field has been the rapid expansion and adoption of *semantic annotators* (Jovanović et al., 2014) - services that take unstructured text as input and annotate/tag it with LOD concepts (i.e., entities defined in LOD knowledge bases such as DBpedia, Wikidata,[21] and Yago). The latter are general, cross-domain knowledge bases storing Wikipedia-like knowledge in well-structured formats with explicitly defined semantics. Several of these LD annotators offer interfaces (APIs) that target the extraction of various types of concepts, such as named entities (e.g., people and places), domain concepts (e.g., protein, gene), and themes or keywords, though the diversity of possible annotations is continuously expanding. Examples of these annotators, both from academia and industry, include DBpedia Spotlight,[22] AlchemyAPI,[23] and TagMe.[24]

Given the plethora of unstructured texts from formal courses, MOOCs, and social media, the capacity of such annotators to produce explicit semantic representations of text makes them valuable for various analytic services. However, very few research works have yet leveraged the power of semantic annotation for learning analytics. Recent research by Joksimović et al. (2015) uses a mixed-method approach for discourse analytics in a cMOOC based on LD and social network analysis (SNA). The aim of the study was to explore the main topics emerging from learners' posts within various social media (i.e., Facebook, Twitter, and blogs) and to analyze how those topics evolve throughout the course (Joksimović et al., 2015). Instead of relying on some of the commonly used topic modelling algorithms (e.g., latent Dirichlet allocation [LDA]), the researchers utilized tools for automated concept extraction (i.e., semantic annotators) along with SNA to identify emerging topics (groups of concepts). Specifically, for each week of the course, concepts were extracted from the posts generated in each of the media analyzed. Further, the authors created graphs based on the co-occurrence of concepts within a single post. Finally, the authors applied modularity algorithm for community detection (Newman, 2006) in order to identify the most prominent groups of concepts (i.e., latent topics). The main advantage of such an approach, over "traditional" topic-modelling algorithms, is possibility to extract compound words (e.g., "complex adaptive systems") that are further linked to knowledge bases (e.g., DBpedia), allowing for easier interpretation of the extracted topics.

Analysis of Scientific Publications in the LA/EDM Field

Another application domain powered by LD and related to the educational context is semantic publishing (e.g., releasing library catalogues as LD) and meta-analysis of scientific publications. In fact, one of the main successes of LD technologies has been their early adoption by various content publishers such as BNF[25] and scientific-based publishing initiatives such as DBLP.[26] This has led to a plethora of LOD vocabularies and datasets related to scientific publications. These datasets provide grounds for various scientometric computations that identify trending topics, influencing researchers, and describe the research community at large (Mirriahi, Gašević, Dawson, & Long, 2014; Ochoa, Suthers, Verbert, & Duval, 2014). They also directly help professionals (researchers, students, librarians, course producers) from the educational sector to locate relevant information.

In the LA/EDM domain, the Learning Analytics and Knowledge (LAK) Dataset (Taibi & Dietze, 2013) represents a corpus of publications from the LA/EDM communities. The LAK Dataset contains both the publications' content and metadata (e.g., keywords, authors, conference). It represents a data integration effort as it relies on various established LOD vocabularies and constitutes a successful application of LD technologies. The analysis of the LAK Dataset has been encouraged since 2013 through the annual LAK Data Challenge, whose goal was to foster research and analytics on the LA/EDM publications. This dataset has been further exploited for the development of data analytics and content analysis applications. One particularly valuable application is the identification of topics and relations between topics in the dataset, per year, per community (LA versus EDM), per publication, and overall. For example, the work of Zouaq, Joksimović, and Gašević (2013) employed ontology learning techniques on the LAK Dataset to identify salient topics and relationships between them. Other techniques applied for discovering topics include latent Dirichlet allocation (LDA; Sharkey & Ansari, 2014) and clustering (Scheffel, Niemann, Leon Rojas, Drachsler, & Specht, 2014). While these approaches offered a text-based content analysis, other works went further in their data integration efforts by relying on the LOD knowledge bases (e.g., DBpedia) and semantic annotators to identify topics of interest. For example, Milikić, Krcadinac, Jovanović, Brankov, and Keca (2013) and Nunes, Fetahu, and Casanova (2013) relied on TagMe and DBpedia Spotlight services, respectively,

[21] https://www.wikidata.org/
[22] https://github.com/dbpedia-spotlight/dbpedia-spotlight/wiki
[23] http://www.alchemyapi.com/
[24] https://tagme.d4science.org/tagme/
[25] http://www.bnf.fr/en/tools/a.welcome_to_the_bnf.html
[26] http://datahub.io/dataset/l3s-dblp

to identify topics and named entities in publications. The benefit of LD in this case was highlighted by 1) the ability to enrich the dataset with LOD concepts, keywords, and themes, and 2) the ability to develop advanced services such as potential collaborator detection (Hu et al., 2014), dataset recommendations, or more general semantic searches (Nunes et al., 2013).

Interpretation of Data Mining Results

Several research works have provided insights, patterns, and predictive models by analyzing learners' interaction and discussion data (e.g., identifying the link between learners' discourse and position and their academic performance (Dowell et al., 2015) or course registration data (d'Aquin & Jay, 2013). However, most of these analyses remain limited to a closed or silo dataset, and are often hard to interpret on large datasets.

In general, pattern discovery in LA/EDM requires a model and a human analyst for the meaningful interpretation of results according to several dimensions (e.g., topics, student characteristics, learning environments, etc.) (d'Aquin & Jay, 2013). The work by d'Aquin & Jay (2013) provides new insights into the usefulness of LD for enriching and contextualizing patterns discovered during the data-mining process. In particular, they propose annotating the discovered patterns with LD URIs so that these patterns can be further enriched with existing datasets to facilitate interpretation. The authors illustrate the idea by a case study of student enrollment in course modules across time. They extract frequent course sequences and enrich them by associating them, via course URIs, with course descriptions, i.e., a set of properties describing the course. The (chain of) properties provide(s) analytical dimensions that are exploited in a lattice-based classification (e.g., the common subjects of frequent course sequences) and as a navigational structure. As illustrated in this case study, LD can help discover new analytical dimensions by linking the discovered patterns to external knowledge bases and exploiting LOD semantic links to infer new knowledge. This is especially relevant in multidisciplinary research where various factors can contribute to a pattern or phenomenon. Given the complexity of learning behaviours, one can imagine the utility of having this support in the interpretation of LA/EDM results.

DISCUSSION AND OUTLOOK

The overall analytical approach to learning experience requires state-of-the-art data management techniques for the collection, management, querying, combination, and enrichment of learning data. The concept and technologies of LD — the latter based on W3C standards (RDF, SPARQL) — have the potential to contribute to all these aspects of data management. First, one of the primary objectives behind LD technologies is to make the data easily processable and reusable, for a variety of purposes, while preserving and leveraging the semantics of the data. Second, LD allows for a decentralized approach to data management by enabling the seamless combination and querying of various datasets. Third, large-scale knowledge bases available as linked open data on the Web provide grounds for a variety of services relevant for the analytic process; e.g., semantic annotators for content analysis and enrichment. Fourth, data exposed as LD on the Web can provide on-demand (just-in-time) data/knowledge input required in different phases of the analytic process, as this knowledge cannot be always fully anticipated in advance. Potential benefits also include representing the resulting analytics in a semantic-rich format so that the results could be exchanged among applications and communicated to interested parties (educators, students) in different manners, depending on needs and preferences (e.g., different visual or narrative forms). Moreover, through its inference capabilities over multiple data sources, originating in semantic-rich representation of data items and their mutual relationships, LD-based methods could be a relevant addition to the existing analytical methods for discovering themes and topics in textual content. More generally, while statistical and machine-learning methods are widespread in the LA/EDM community, other kinds of data analysis methods and techniques — those based on explicitly defined semantics of the data — and open knowledge resources (especially open, Web-based knowledge) can make the traditional analytical approaches even more powerful. Some of the potential enrichments provided by LD include semantic vector-based models (e.g., bags of concepts instead of bags of words), semantic-rich social network analysis with explicitly defined semantics for edges and nodes, or recommendations based on semantic similarity measures.

Finally, LD technologies can be useful in dealing with the heterogeneity of learning environments and social media platforms. In particular, one can query and assemble various datasets that do not share a common schema. This aspect in itself represents a more flexible and practical approach than previous approaches that required compliance to a common model/schema.

However, there are also several challenges related to the use of LD in terms of the following:

1. **Quality:** The quality of the LOD datasets is a concern (Kontokostas et al., 2014), and linking learning resources and traces to external datasets and knowledge bases might introduce noisy data. Although there are some initiatives for data

cleaning, this issue is far from being resolved.

2. **Alignment:** Besides the use of common Web URIs among schemas, there is often a need to semantically align vocabularies and models, which is a challenging task. Current alignment approaches are often based on syntactic matching, which does not deal well with ambiguities. One way to mitigate the alignment issue is to be aware and re-use major LD vocabularies[27] whenever possible (e.g., *foaf:name* is a property depicting the name of a person in the FOAF specification and could be used instead of creating a new property);
3. **Privacy:** Data within MOOCs and learning platforms is often siloed for privacy reasons. Merging information between learning and social platforms would require, for example, that learners grant access to their data and provide log-in information for the different services they use for learning.

Despite the challenges indicated above - and given the use of LOD datasets and knowledge bases in some major initiatives such as Google knowledge graph or Facebook graph search and their increasing adoption in educational institutions - LD is a promising technological backbone for today's learning platforms. It also provides a useful formalism for facilitating the overall learning analytic process, from raw data collection and storage, to data exploitation and enrichment, to interpretation of the analytics results.

REFERENCES

Belleau, F., Nolin, M.-A., Tourigny, N., Rigault, P., & Morissette, J. (2008). Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. *Journal of Biomedical Informatics*, 41(5), 706–716.

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data - the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. Preprint retrieved from http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. International *Journal of Technology Enhanced Learning*, 4(5–6), 318–331.

Cooper, A. R. (2013). Learning analytics interoperability: A survey of current literature and candidate standards. http://blogs.cetis.ac.uk/adam/2013/05/03/learning-analytics-interoperability

d'Aquin, M., & Jay, N. (2013). Interpreting data mining results with linked data for learning analytics: Motivation, case study and directions. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 155–164). New York: ACM.

d'Aquin, M., Adamou, A., & Dietze, S. (2013, May). Assessing the educational linked data landscape. *Proceedings of the 5th Annual ACM Web Science Conference* (WebSci '13), 2–4 May 2013, Paris, France (pp. 43–46). New York: ACM.

De Nies, T., Salliau, F., Verborgh, R., Mannens, E., & Van de Walle, R. (2015, May). TinCan2PROV: Exposing interoperable provenance of learning processes through experience API logs. *Proceedings of the 24th International Conference on World Wide Web* (WWW '15), 18–22 May 2015, Florence, Italy (pp. 689–694). New York: ACM.

Desmarais, M. C., & Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1–2), 9–38.

Dietze, S., Drachsler, H., & Giordano, D. (2014). A survey on linked data and the social web as facilitators for TEL recommender systems. In N. Manouselis, K. Verbert, H. Drachsler, & O. C. Santos (Eds.), *Recommender Systems for Technology Enhanced Learning* (pp. 47–75). New York: Springer.

Dietze, S., Sanchez-Alonso, S., Ebner, H., Qing Yu, H., Giordano, D., Marenzi, I., & Nunes, B. P. (2013). Interlinking educational resources and the web of data: A survey of challenges and approaches. *Program*, 47(1), 60–91.

Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., & Taibi, D. (2012). Linked education: Interlinking educational resources and the web of data. *Proceedings of the 27th Annual ACM Symposium on Applied Computing* (SAC 2012), 26–30 March 2012, Riva (Trento), Italy (pp. 366–371). New York: ACM.

[27] http://lov.okfn.org/dataset/lov/

Dowell, N. M., Skrypnyk, O., Joksimović, S., Graesser, A. C., Dawson, S., Gasevic, D., Hennis, T. A., de Vries, P., & Kovanović, V. (2015). Modeling learners' social centrality and performance through language and discourse. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Education Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 250–257). International Educational Data Mining Society. http://files.eric.ed.gov/fulltext/ED560532.pdf

Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (LAK '11), 27 February–1 March 2011, Banff, AB, Canada (pp. 9–17). New York: ACM.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71.

Groth, P., Loizou, A., Gray, A. J., Goble, C., Harland, L., & Pettifer, S. (2014). API-centric linked data integration: The open PHACTS discovery platform case study. *Web Semantics: Science, Services and Agents on the World Wide Web*, 29, 12–18.

Heath, T., & Bizer, C. (2011). Linked data: Evolving the web into a global data space. *Synthesis lectures on the semantic web: Theory and technology*, 1(1), 1–136. Morgan & Claypool.

Hu, Y., McKenzie, G., Yang, J. A., Gao, S., Abdalla, A., & Janowicz, K. (2014). A linked-data-driven web portal for learning analytics: Data enrichment, interactive visualization, and knowledge discovery. In LAK Workshops. http://geog.ucsb.edu/~jano/LAK2014.pdf

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gašević, D., & Hatala, M. (2015). What do cMOOC participants talk about in social media? A topic analysis of discourse in a cMOOC. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March, Poughkeepsie, NY, USA (pp. 156–165). New York: ACM.

Jovanović, J., Bagheri, E., Cuzzola, J., Gašević, D., Jeremic, Z., & Bashash, R. (2014). Automated semantic annotation of textual content. *IEEE IT Professional*, *16*(6), 38–46.

Kagemann, S., & Bansal, S. (2015). MOOCLink: Building and utilizing linked data from massive open online courses. *Proceedings of the 9th IEEE International Conference on Semantic Computing* (IEEE ICSC 2015), 7–9 February 2015, Anaheim, California, USA (pp. 373–380). IEEE.

Kessler, C., d'Aquin, M., & Dietze, S. (2013). Linked data for science and education. *Journal of Semantic Web*, 4(1), 1–2.

Kontokostas, D., Westphal, P., Auer, S., Hellmann, S., Lehmann, J., Cornelissen, R., & Zaveri, A. (2014). Test-driven evaluation of linked data quality. *Proceedings of the 23rd International Conference on World Wide Web* (WWW '14), 7–11 April 2014, Seoul, Republic of Korea (pp. 747–758). New York: ACM.

Lausch, A., Schmidt, A., & Tischendorf, L. (2015). Data mining and linked open data: New perspectives for data analysis in environmental research. *Ecological Modelling*, 295, 5–17.

Maturana, R. A., Alvarado, M. E., López-Sola, S., Ibáñez, M. J., & Elósegui, L. R. (2013). Linked data based applications for learning analytics research: Faceted searches, enriched contexts, graph browsing and dynamic graphic visualisation of data. LAK Data Challenge. http://ceur-ws.org/Vol-974/lakdatachallenge2013_03.pdf

Milikić, N., Krcadinac, U., Jovanović, J., Brankov, B., & Keca, S. (2013). Paperista: Visual exploration of semantically annotated research papers. LAK Data Challenge. http://ceur-ws.org/Vol-974/lakdatachallenge2013_04.pdf

Mirriahi, N., Gašević, D., Dawson, S., & Long, P. D. (2014). Scientometrics as an important tool for the growth of the field of learning analytics. *Journal of Learning Analytics*, *1*(2), 1–4.

Muñoz-Merino, P. J., Pardo, A., Kloos, C. D., Muñoz-Organero, M., Wolpers, M., Katja, K., & Friedrich, M. (2010). CAM in the semantic web world. In A. Paschke, N. Henze, & T. Pellegrini (Eds.), *Proceedings of the 6th International Conference on Semantic Systems* (I-Semantics '10), 1–3 September 2010, Graz, Austria. New York: ACM. doi:10.1145/1839707.1839737

Newman, M. E. (2006). Modularity and community structure in networks, *Proceedings of the National Academy of Sciences*, *103*(23), 8577–8582.

Niemann, K., Wolpers, M., Stoitsis, G., Chinis, G., & Manouselis, N. (2013). Aggregating social and usage datasets for learning analytics: Data-oriented challenges. *Proceedings of the 3rd International Conference on Learning Analytics and Knowledge* (LAK '13), 8–12 April 2013, Leuven, Belgium (pp. 245–249). New York: ACM.

Nunes, B. P., Fetahu, B., & Casanova, M. A. (2013). Cite4Me: Semantic retrieval and analysis of scientific publications. LAK Data Challenge, 974. http://ceur-ws.org/Vol-974/lakdatachallenge2013_06.pdf

Ochoa, X., Suthers, D., Verbert, K., & Duval, E. (2014). Analysis and reflections on the third learning analytics and knowledge conference (LAK 2013). *Journal of Learning Analytics*, *1*(2), 5–22.

Piedra, N., Chicaiza, J. A., López, J., & Tovar, E. (2014). An architecture based on linked data technologies for the integration and reuse of OER in MOOCs context. *Open Praxis*, 6(2), 171–187.

Santos, J. L., Verbert, K., Klerkx, J., Duval, E., Charleer, S., & Ternier, S. (2015). Tracking data in open learning environments. *Journal of Universal Computer Science*, *21*(7), 976–996.

Scheffel, M., Niemann, K., Leon Rojas, S., Drachsler, H., & Specht, M. (2014). Spiral me to the core: Getting a visual grasp on text corpora through clusters and keywords. LAK Data Challenge. http://ceur-ws.org/Vol-1137/lakdatachallenge2014_submission_3.pdf

Schmitz, H. C., Wolpers, M., Kirschenmann, U., & Niemann, K. (2011). Contextualized attention metadata. In C. Roda (Ed.), *Human attention in digital environments* (pp. 186–209). New York: Cambridge University Press.

Sharkey, M., & Ansari, M. (2014). Deconstruct and reconstruct: Using topic modeling on an analytics corpus. LAK Data Challenge. http://ceur-ws.org/Vol-1137/lakdatachallenge2014_submission_1.pdf

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1), 3–10. http://itdl.org/Journal/Jan_05/article01.htm

Softic, S., De Vocht, L., Taraghi, B., Ebner, M., Mannens, E., & De Walle, R. V. (2014). Leveraging learning analytics in a personal learning environment using linked data. *Bulletin of the IEEE Technical Committee on Learning Technology*, 16(4), 10–13.

Taibi, D., & Dietze, S. (2013), Fostering analytics on learning analytics research: The LAK dataset. LAK Data Challenge, 974. http://ceur-ws.org/Vol-974/lakdatachallenge2013_preface.pdf

Zablith, F. (2015). Interconnecting and enriching higher education programs using linked data. *Proceedings of the 24th International Conference on World Wide Web* (WWW '15), 18–22 May 2015, Florence, Italy (pp. 711–716). New York: ACM.

Zotou, M., Papantoniou, A., Kremer, K., Peristeras, V., & Tambouris, E. (2014). Implementing "rethinking education": Matching skills profiles with open courses through linked open data technologies. *Bulletin of the IEEE Technical Committee on Learning Technology*, 16(4), 18–21.

Zouaq, A., Joksimović, S., & Gašević, D. (2013). Ontology learning to analyze research trends in learning analytics publications. LAK Data Challenge. http://ceur-ws.org/Vol-974/lakdatachallenge2013_08.pdf