RESEARCH & PRACTICE IN ASSESSMENT

Special Issue: Big Data & Learning Analytics

VOLUME NINE | WINTER 2014
www.RPAjournal.com
ISSN# 2161-4210

RPA

A PUBLICATION OF THE VIRGINIA ASSESSMENT GROUP

Editor's Farewell

This special issue on Big Data & Learning Analytics marks the final issue for me as Editor of RPA. In 2010, I began with a vision to overhaul the publication with a tripartite focus on developing: (a) Disciplinary Convergence, (b) Scholarly Quality, and (c) Visual Aesthetics. With these foci, the journal sought to press the field of educational assessment to continue to innovate in ways that remained current with the changing landscape of higher education. Specifically, to think beyond the dominant frameworks of the profession (psychometrics, rubrics, and standards), and to persistently engage the immanent contextual factors facing the field, namely those in the social, cultural, historical, political, philosophical, economic, and technological spheres. Four years and eight issues later, I feel the scholars and practitioners listed on the previous page have admirably collaborated in a manner such that RPA has successfully navigated beyond adolescence.

I would like to thank the current and past board members of the Virginia Assessment Group for their confidence in my stewardship of the publication for four years. During this time I was provided with the resources and the freedom to transform the journal to its present state. From the beginning, two persons shouldered the bi–annual production weight with me, Alysha Clark (Editorial Assistant) and Patrice Brown (Graphic Designer). They provided *countless* hours of service, making the journey possible. For the myriad persons who worked with me during this time, thank you kindly for tolerating my persistence and determination to forge a new space in the assessment discourse. With the publication of this issue, Katie Busby, Assistant Provost for Assessment and Institutional Research at Tulane University, assumes the editorial leadership of RPA. Between her commanding knowledge of the field and the faithful contributions of service made by members of the RPA Editorial and Review Boards, I am confident the best days for the journal are yet to come.

Regards,

Joshua Brown

RPA Editor, 2010–2014

RESEARCH & PRACTICE IN ASSESSMENT

The goal of *Research & Practice in Assessment* is to serve the assessment community as an online journal focusing on higher education assessment. It is dedicated to the advancement of scholarly discussion amongst researchers and practitioners in this evolving field. The journal originated from the Board of the Virginia Assessment Group, one of the oldest continuing professional higher education assessment organizations in the United States. *Research & Practice in Assessment* is a peer-reviewed publication that uses a double-blind review process. Approximately forty percent of submissions are accepted for issues that are published twice annually. *Research & Practice in Assessment* is listed in Cabell's Directory and indexed by EBSCO, Gale, and ProQuest.

Published by:
VIRGINIA ASSESSMENT GROUP | www.virginiaassessment.org

Publication Design by Patrice Brown |

TABLE OF CONTENTS

FROM THE EDITOR

Flatlands & Frontiers

"Escaping this flatland is the essential task of envisioning information – for all the interesting worlds (physical, biological, imaginary, human) that we seek to understand are inevitably and happily multivariate in nature. Not flatlands."

Edward R. Tufte, *Envisioning Information,* 1990

The world of higher education is multivariate. It is a multidimensional and complex realm we seek to further understand. And yet, the world portrayed in our assessments often focuses on a single dimension. They are flatlands. These reproductions are crafted using data, rubrics, psychometrics, standards, and "cycles." However, while we were yet producing portraits of the higher education landscape, a new data type emerged that was not one-dimensional. Someone named it with an adjective – big.

The publication of this issue makes no claim that big data or learning analytics are a panacea for the multivariate world of higher education. Persons should not pretend that big data will solve what policy analysts call "wicked problems," those utterly complex educational and social ills. Rather, this issue seeks to begin a collective debate about the extent to which big data and learning analytics might play a role in higher education assessment. As such, the works in this issue commence with the essential tasks necessary for interrogating an emerging body of knowledge: they examine assumptions, operationalize terms, suggest new metrics, compare educational sectors, consider implications for policy, and scrutinize professional ethics. I have previously argued in this column that in order to move beyond the flatlands of assessment the disciplines must be converged within the assessment discourse. This special issue is no different - it seeks to converge the learning analytics and assessment literatures.

The pieces in this issue have been arranged to provide a natural progression on the topic for the reader. The volume opens with an article by Candace Thille et al. that provides a definition of big data and examines how assessment processes with large-scale data will be different from those without it. Emphasizing a "wicked" problem in a complex system, Leah Macfadyen, Shane Dawson, Abelardo Pardo & Dragan Gaševic̀ offer a policy framework for navigating the tension between assessment-for-accountability and assessment-for-learning. Matters pertaining to various analytics are then given attention beginning with Una-May O'Reilly & Kalyan Veeramachaneni who describe an agenda for developing technology that enables MOOC analytics. Ryan Baker & Albert Corbett then consider how an emphasis on robust learning might advance the focus of assessments from single to multiple domains. Following this, Maarten de Laat & Fleur Prinsen introduce social learning analytics as an instrument in formative assessment practices. The final two articles offer innovative systems presently being used in organizations to strengthen student success through persistence and retention. In the first, Tristan Denley highlights how closing the information gap impacts the educational achievement gap for low income and minority students. Mark Milliron, Laura Malcolm & David Kil use insight and action analytics to produce predictive flow models of student progression and completion across three diverse organizations.

Book reviews for this volume were strategically chosen to provide readers with a sample of present works on big data. Aiden & Michel's accessible work based on the Google Ngram Viewer, *Uncharted: Big data as a lens on human culture* is reviewed by Carolyn Penstein Rose. Fabio Rojas then engages Lane's *Building a Smarter University: Big data, innovation, and analytics*, suggesting this may be an important volume for university administrators. Finally, drawing parallels from the K-12 sector, Karly Sarita Ford reviews Piety's book *Assessing the Educational Data Movement*. The end of the issue asks readers to give consideration to the myriad subjects of big data. Here, Mitchell Stevens poignantly ask us to consider the legal, political and ethical questions of big data collection. He highlights the heroic efforts of the scholars and scientists at the Asilomar Convention, which yielded six principles to inform the navigation of this uncertain terrain.

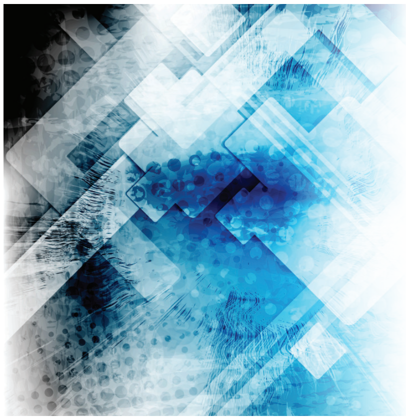
While the flatlands offer a rich and fertile soil, I am not content simply looking afar at the majesty of the mountains. The teacher that resides deep within me wants to use learning analytics to venture beyond the plains, to scale the summit of the multivariate. I want to reside on the frontier of the discipline, knowing that I will not meet my fate in the infinite cycle of assessment. As you engage the pages herein, give consideration as to how the frontiers of the discipline may continually be explored.[1]

Regards,

Joshua Brown

Liberty University

[1] The framing of this column was influenced by the scholarship of Edward R. Tufte (1990) and Emma Uprichard (2014).



Abstract

The article addresses the question of how the assessment process with large–scale data derived from online learning environments will be different from the assessment process without it. Following an explanation of big data and how it is different from previously available learner data, we describe three notable features that characterize assessment with big data and provide three case studies that exemplify the potential of these features. The three case studies are set in different kinds of online learning environments: an online environment with interactive exercises and intelligent tutoring, an online programming practice environment, and a massive open online course (MOOC). Every interaction in online environments can be recorded and, thereby, offer an unprecedented amount of data about the processes of learning. We argue that big data enriches the assessment process by enabling the continuous diagnosis of learners' knowledge and related states, and by promoting learning through targeted feedback.

AUTHORS

Candace Thille, Ed.D.

Emily Schneider, B.A.

René F. Kizilcec, B.A.

Christopher Piech, M.S.

Sherif A. Halawa, M.Sc.

Daniel K. Greene, B.A.
Stanford University

The Future of Data–Enriched Assessment

A fundamental goal of education is to equip people with the knowledge and skills that enable them to think critically and solve complex problems. The process of quantifying the degree to which people have acquired such knowledge and skills is at the heart of assessment. Over the last decades, large–scale assessment of knowledge has become increasingly standardized, primarily to provide policy and other decision makers with clearer signals on the effectiveness of educational institutions and practices (Shavelson, 2007). Yet the merits of effective assessment extend far beyond informing policy decisions: instructors can gain valuable insights into the effectiveness of their instructional methods and learners receive feedback on their learning approach and overall progress. In providing an opportunity to apply the acquired knowledge and skills with subsequent feedback, assessment can promote learning if designed appropriately (Black & Williams, 1998; Gikandia, Morrowa, & Davisa, 2011; Roediger & Karpicke, 2006).

Education is becoming ever more augmented by technology to create new ways of interacting with educational content and communicating with instructors and peers. A number of promising technologies fall under the broad category of online learning environments, which rely on digital, networked systems but vary substantially in the features they provide to instructors and learners. Some such environments attempt to provide a holistic learning experience by integrating instruction, assessment, and social interaction. Other environments serve as a complementary resource to augment an in–person learning experience. In this paper, we present three case studies, which are set in different kinds of online learning environments: an online environment with interactive exercises and intelligent tutoring, an online programming practice environment, and a massive open online course (MOOC). The latter is an online learning environment in which thousands

CORRESPONDENCE

Email
cthille@stanford.edu

of people worldwide can learn about a given topic from lecture videos, quiz questions, longer assignments, and discussions with peers on a forum, to name but a few of the many forms of interaction that can occur in these environments (Kizilcec, Piech, & Schneider, 2013). Similar to non–educational online content providers, every interaction in these environments can be recorded and, thereby, offer an unprecedented amount of data about the processes of learning.

Online learning environments hold the potential to better support learning and to create opportunities for novel forms of assessment. The question we address in this article is: how will the assessment process with large–scale data derived from online learning environments be different from the assessment process without it? To address this question, we first explain our definition of big data, and how we believe it is different from previously available learner data. We then present three notable features that characterize assessment with big data and provide three case studies that exemplify the potential of these features. We argue that big data enriches the assessment process by enabling the continuous diagnosis of learners' knowledge and related states, and by promoting learning through targeted feedback.

Big Data

How will the assessment process with large–scale data derived from online learning environments be different from the assessment process without it?

Big data, in the context of assessment, is learner data that is deep as well as broad.[1] Large amounts of data can occur not only across many learners (broad between–learner data), but also within individual learners (deep within–learner data). Moreover, the depth of data is determined not only by the raw amount of data on a given learner, but also by the availability of contextual information that adds semantic meaning to within–learner data. Clickstream data is a good example of big data that tends to fall short of providing meaningful information in the context of assessing learning (cf. Case Study 1), although it may be sufficiently deep for assessing persistence (cf. Case Study 3). Therefore, the dimensionality of big data depends fundamentally on the object of assessment. More importantly, the converse is also true: new forms of data–enriched assessment require collecting deeper and broader data in order to gain insight into the new object of assessment.

Large–scale standardized tests, for instance, are broad but not deep; they yield large amounts of data consisting of test scores for thousands of learners with the primary focus of providing comparisons across learners, but which provide relatively little information about each individual learner. In contrast, a virtual reality learning experience (e.g., a mathematics lesson in a virtual classroom) can track learners' body positions to generate a substantial amount of behavioral and other information, but only for a small number of learners. Data–enriched assessment in appropriately instrumented online learning environments can, for a large number of learners, provide insights into each individual learner's problem–solving processes, strategic learning choices, misconceptions, and other idiosyncratic aspects of performance. In practice, this typically implies that information about learner performance is plentiful enough to gain new insights by applying modern data mining and machine learning methods (Romero, Ventura, Pechenizkiy, & Baker, 2011), such as hidden Markov modeling (cf. Case Study 1), probabilistic graphical modeling (cf. Case Study 2), or natural language processing methods (cf. Baker & Corbett, 2014).

Previously available data in assessment have been large in one of the two dimensions, but rarely before have education researchers been in a position to collect large amounts of data on both dimensions at once. The promise of big data in online learning environments is that capturing semantically meaningful information both across and within learners provides a powerful basis for assessing and supporting learners.

Elements of Data–Enriched Assessment

Deep and broad learner data in an interactive online learning environment can enable assessment tasks that are continuous, feedback–oriented, and multifaceted.

Continuous. In an online learning environment, an individual's learning process can be continually observed: the steps in solving a math problem, the chemicals combined on a virtual lab bench, and the learner's contributions to a discussion forum are all captured by the system. Interactions with learning resources, with peers, or with the instructor each contain

[1] A technical definition of big data focuses on the technological constrains that occur during computation, and which tend to require distributed processing and approximations instead of exact computations.

evidence about the concepts and skills over which the learner currently has command. There is no need to distinguish between learning activities and moments of assessment. Instead, a model of the learner's knowledge state is continually assessed and updated – as are models of other facets of the learner, as described below. This enables learning to be modeled as an ongoing process rather than as a set of discrete snapshots over time.

Feedback–oriented. Feedback is central to an assessment process that is designed to support learning. Well–designed feedback presents the learners' current state, along with enough information to make a choice about the appropriate next action. Feedback can be provided directly to the learner, to an instructor, or to the system (e.g., an adaptive test or an intelligent tutor). Providing learners with the choice of when to receive feedback and an opportunity to reflect on feedback may have additional benefits for developing metacognitive competencies. Drawing on prior work on the relative benefits of different types of feedback for learners with particular characteristics, online learning environments can also provide personalized feedback. For instance, based on a design principle proposed by Shute (2008) in a review of the feedback literature, the system could offer direct hints to low–achieving learners and reflection prompts to higher–achieving learners.

The effective presentation of feedback in online learning environments poses an interesting design challenge. Graphs, maps, and other information visualization techniques can be used to represent learner progress through the multiple concepts and competencies that learners are attempting to master. The information visualization community has developed an increasingly sophisticated visual language for representing complex datasets (e.g., Ware, 2013), and the efficacy of particular visualization strategies for supporting learners and instructors is a fruitful area for future research.

Multifaceted. Learners' abilities to learn from resources or interactions with others is influenced by factors beyond their current knowledge state. There are many reasons that a learner may start a task, struggle with it, or complete it successfully. Detecting these factors can contextualize observations about cognitive competencies, which provides the system or an instructor with additional information to target feedback or an intervention. The learner's life context is an important facet for developing deeper understanding of the learner's experience (cf. Case Study 3). Affective state – the learner's mood or emotions – can also have an impact on the learning processes (cf. Baker & Corbett, 2014), as can interpersonal competencies, such as the ability to communicate and collaborate effectively with others (De Laat & Prinsen, 2014).

Big data, in the context of assessment, is learner data that is deep as well as broad. Large amounts of data can occur not only across many learners (broad between–learner data), but also within individual learners (deep within–learner data).

Other critical facets of the learner include self–regulation – a learner's awareness and effective application of study strategies (Zimmerman, 1990); goal orientation – a learner's purpose in engaging with the learning activity (Pintrich, 2003); and mindset – a learner's beliefs about whether intelligence is fixed or malleable (Dweck, 2006). In addition, a rich history of research in social and educational psychology highlights the impact of learners' attributions of social cues in their environment (Cohen & Sherman, 2014; Steele, 1997), for example, whether a learner experiences a sense of social belonging in an environment (Walton & Cohen, 2011). Each of these intrapersonal, affective, contextual, and interpersonal states can be included in a model as latent states of the learner or directly reported features. Complex multifaceted models are enabled by big data and can advance research on the impact of each of these factors on learning.

The multiple facets of a learner translate into key competencies for individuals to be productive and resilient in future educational and professional settings. Explicitly assessing these competencies as desired outcomes of learning can inform the design of learning environments to support their development and thereby better serve learners for the long term.

Case Studies

In the following case studies, we draw on our work in three online learning environments to describe multiple approaches to data–enriched assessment. In each case study, learner data is deep because the learner is observed continuously, and broad as a result of the number of learners who engage with the online learning environment. Additional data dimensionality is added by specifying the relationship of learner activities to the concepts requisite for successful task engagement (Case Study 1) and to the appropriate next steps in a problem–solving process (Case Study 2). This specification, or "expert labeling," can occur in advance of developing an initial model or in the process of refining a learner model. Regardless of variations in the object of assessment or the timing of expert labeling, each case study uses machine learning techniques to develop or refine a learner state model.

In Case Study 1, the Open Learning Initiative, the assessment tasks are designed and embedded within the learning process. Data collected on learner performance on assessment tasks are used to diagnose the knowledge state of the learner and give feedback in real time and to refine underlying models. In Case Study 2, learners engage in open–ended software programming tasks, and assessment is focused on the processes of problem solving. Moreover, patterns in these processes are used to automatically generate suggestions for future learners who are struggling with the task. Case Study 3, focused on MOOCs, addresses the challenge of assigning meaning to learner activities that are outside of problem solving, such as forum interactions and video watching habits.

An intelligent tutor is a computer program whose design is based on cognitive principles and whose interaction with learners is based on that of a good human tutor, making comments when the learner errs, answering questions about what to do next, and maintaining a low profile when the learner is performing well.

Case study 1: The open learning initiative (OLI). *Open Learning Initiative (OLI)* at Stanford University and Carnegie Mellon University is a grant funded open educational resources initiative. Data have been collected from over 100,000 learners that have participated in an OLI course for credit at academic institutions of all Carnegie Classifications and from over 1,000,000 learners that have engaged in one of the free and open versions of an OLI course.

OLI courses comprise sequences of expository material such as text, demonstration videos and worked examples interspersed with interactive activities such as simulations, multiple choice and short answer questions, and virtual laboratories that encourage flexible and authentic exploration. Perhaps the most salient feature of OLI course design is found in the intelligent tutors embedded within the learning activities throughout the courses. An intelligent tutor is a computer program whose design is based on cognitive principles and whose interaction with learners is based on that of a good human tutor, making comments when the learner errs, answering questions about what to do next, and maintaining a low profile when the learner is performing well. The tutors in OLI courses provide the learner tailored feedback to individual responses, and they produce data.

OLI learning environments and data systems have been designed to yield data that inform explanatory models of a student's learning that support course improvement, instructor insight, learner feedback, and the basic science of learning. Modern online learning environments can collect massive amounts of learner interaction data; however, the insights into learning that can be gleaned from that data are limited by the type of interaction that is observable and by the semantic tagging (or lack of tagging) of the data generated by the interaction. Many MOOC platforms and traditional learning management systems collect clickstream data that can measure frequency and timing of learner log–ins, correctness (or incorrectness) of learner responses, learner use of resources, and learner participation in forums. While such clickstream data may be used to predict which learners are likely to complete the course, they do not explain if or how learning is occurring.

In OLI, the learning data are organized by learning objective. Learning objectives identify what a learner should be able to do or demonstrate they know by the end of the learning experience. Each learning objective comprises one or more skills. Skills break down the learning objective into more specific cognitive processes.

The course design process starts with the articulation of the learning objectives and skills. During the design of the course, the opportunities for learner action (e.g., answering a question, taking a step in a multi–step task, acting on a simulation) in an interactive activity are associated with the learning objectives and skills. The relationships among learning objectives, skills and learning activities are fully many–to–many: each learning objective may have one or more component skills, each skill may contribute to one or more learning objectives, each skill may be assessed by one or more steps in a task, each task step may assess one or more skills. Typical OLI courses comprise about 30 to 50 learning objectives and 100 to 1,000 skills.

Teams of faculty domain experts, learning scientists, human–computer interaction experts, assessment experts, and software engineers work collaboratively to develop the OLI courses and a parameterized model that predicts learner mastery of component skills. Skills are ranked as easy, moderate, or difficult based on perceived complexity. Initially, the labels are based on an analysis of the domain and on the expert's prior teaching experience. The rankings are used to adjust baseline parameters and, during the initial design of the course, the adjustments are heuristic, not empirical. The model associates learner practice with individual skills rather than with larger topics in a domain or activity in the course in general. The underlying theory is that learning is skill specific and it is practice on the specific skill that matters rather than practice in the course in general.

The skill model that the development team has created is considered naïve until it has been validated by data. Machine learning algorithms support learning researchers to improve upon the initial human–generated model by searching for models of learning that produce a better fit to the learner–generated data. The algorithms split and combine existing skills and suggest new skills where appropriate but, to date, a human must supply labels for the changes suggested by the algorithm. The researchers use the data to evaluate the fit of the model and to tune the parameters for the model. The course design team also uses the data to refine the learning activities and the response–level feedback.

While such clickstream data may be used to predict which learners are likely to complete the course, they do not explain if or how learning is occurring.

The skill model serves a number of purposes, including assisting in the iterative course improvement process; measuring, validating and improving the model of learning that underlies each course; and offering information necessary to support learning scientists in making use of OLI datasets for continued research. In the original versions of OLI courses, learning is modeled using a Bayesian hierarchical statistical model with the latent variables of interest, learners' knowledge state, becoming more accurate as more data is accrued about performance on a given skill. Skills are modeled using a multi–state hidden Markov model. The Markov model is hidden because the knowledge states cannot be observed directly; inferences need to be made about which state a learner is in based on the learner's answers to questions. In the original models, individual skills are treated as mathematically independent variables and it is assumed that learning a skill is a one–way process: once a skill is learned, it is not unlearned.

One of the most important uses of the skill model is to support learning analytics for instructors and learners. The OLI system analyzes the learner activity in real time against the skill model. When a learner responds to a question or engages in an OLI activity, the system uses the skill model mapping to identify the skills related to that question or activity. The learning estimates are computed per skill per learner and use simple algorithms with low computational overhead to allow real time updates. Data are aggregated across skills for a given learning objective and reported to instructors and students at that level. It is this real time feedback to instructors and students about mastery of learning objectives that helps guide the instructional and learning process throughout the course.

Case study 2: Code webs. The Code Webs Project is a Stanford machine learning research collaboration to analyze logs of learners completing open ended programming assignments with the intention to (a) uncover new perspectives into individual learner abilities, (b) paint a picture of how learners in general approach problems, and (c) understand how to help learners navigate complex programming assignments.

The project studies logs of learners solving assignments in three courses: The Code.org Hour of Code (Code.org), The Coursera Machine Learning class (ML) and Stanford's Introduction to Computer Science course (CS1). The Code.org and ML courses are both open access online courses, whereas the CS1 is a traditional in–person college course. The data are wide and deep. In each course learners complete a set of challenging programming tasks and each time a learner saves or runs an intermediate solution to a task, an entire snapshot of their current work is recorded. When the learner submits a final answer, or stops working on an assignment, all of the learner's partial solutions are composed into a trajectory. From the three courses, the Code Webs project has compiled trajectories from over 1,000,000 learners.

One of the most generally applicable results of this research has been to demonstrate the tremendous potential towards better assessment that comes from digital logs of how learners work through assignments, as opposed to just the learner's final submission. In future educational settings, the data on how learners develop their homework solutions from start to finish will become more ubiquitous and machine learning techniques applied to this format of data will generate important insights.

The first nugget that can be discovered from learner trajectories is a depiction of how learners, both as a cohort and individually, solve open ended work. In CS1, the Code Webs team instrumented the programming environment that learners used to generate their homework solutions. Using the data gathered, the research team modeled how groups of learners proceed through the assignment, using a Hidden Markov model that involved:

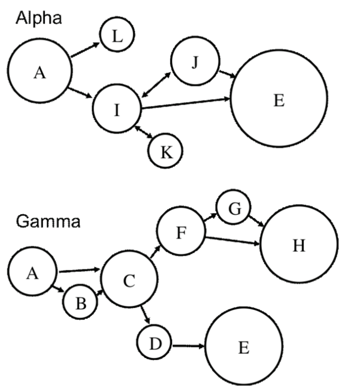
a. Inferring the finite set of high–level states that a partial solution could be in.

b. The transition of probabilities of a learner moving from one state to another.

c. The probability of seeing a specific partial solution given that a learner is in a state.

Once transition patterns for each learner had been fit, we then clustered the transition patterns to produce different prototypical ways that learners approach programming assignments.

In the CS1 dataset we discovered two notable prototypical patterns: A "Gamma" group whose progress is defined by steady work towards the objective and an "Alpha" group in which learners tend to get stuck in states where they would spend a lot of time before moving back to a previous state and then manage to make a large jump to a solution. Figure 1 demonstrates the pattern for a particular assignment in CS1.



State	Description
A	Start state.
B	Places a single beeper.
C	Places a perfect row in all worlds.
D	Places two perfect rows of the comb soln.
E	The comb soln. that works on all worlds.
F	The zigzag soln. that works except one col worlds.
G	The zigzagsoln. that crashes on odd col worlds.
H	The zigzag soln. that works on all worlds.
I	Places one line. Misses a beeper on odd col worlds.
J	Two rows of the zigzag soln. Second row shifted.
K	Repeats first two rows infinitely.
L	Hardcoded to place first row on the 8x8 world.

Figure 1. Visualization of the two prototypical patterns for solving an open ended assignment in CS1. While most learners submitted a correct final solution, how they arrived at their answer was more predictive of miderm score. Only the most popular states and transitions are visualized.

One of the most generally applicable results of this research has been to demonstrate the tremendous potential towards better assessment that comes from digital logs of how learners work through assignments, as opposed to just the learner's final submission.

In CS1, almost all learners turn in working solutions to the class assignments; however on the class midterms and finals, some learners are unable to solve basic tasks. A promising result of this work was that the learners' problem solving patterns on the first assignment were more predictive of midterm grades than were their final scores on the first assignment.

Data logs on learners' solving problems can give insights into how learners are approaching problems and to what extent they understand the material. In addition to finding prototypical patterns, the autonomous process also computes to what extent each learner's progress matches the common patterns, and the overall distribution of the class.

Trajectories can also be used to autonomously learn what learners should do when working on open ended problems. For example, if we observe thousands of past learners who got stuck on the same problem, it seems plausible that we could use the subsequent actions that they took to predict the ideal solution to that problem. To explore this avenue, the Code Webs project team looked at the trajectories from half a million predominantly middle school learners solving the same programming assignments in Code.org's Hour of Code. We devised an experiment where experts generated a strategy of what partial solution a learner should transition to next given their current state and, using trajectory data, learn an algorithm that could recreate the expert strategy.

Surprisingly, many reasonable statistics on learner trajectories are not particularly useful for predicting what expert teachers say is the correct way forward. The partial solution that learners most often transition to after encountering a problem does not tend to correspond with what experts think learners should do. The wisdom of the crowd of learners, as seen from this angle, is not especially wise. However, there are other signals from a large corpus of trajectory data that shed light onto what a learner should do from a current partial solution. One algorithm generates a data–driven approximation of a complete journey from any current state to a solution that it expects would be most common if students were evenly distributed amongst the problem solving space. The first step in the generated journey overwhelmingly agrees with expert labels of how learners should proceed. This algorithm can be applied to logs of learners working on problems for which there are no expert labels, and will produce an intelligent strategy for what learners ought to do.

By modeling how learners progress through an assignment we open up the possibility for data driven feedback on problem solving strategies. By learning a predictor for how experts think a learner should proceed through a project, the process for generating a hint is simplified, both because we know what part of an open ended problem a stuck learner should work on next and we know what change they should make. Since the feedback can be autonomously generated it could be continuously and immediately provided to learners.

In future educational settings, the data on how learners develop their homework solutions from start to finish will become more ubiquitous and machine learning techniques applied to this format of data will generate important insights.

Trajectories seem like a promising medium through which we can leverage large amounts of data to generate better and more scalable assessment for learners that do their work in an instrumented environment. Though this case study was about computer programming, the algorithms used would apply to any trajectories of learner data, given an appropriate representation of partial solutions. While the Code Webs project has made progress towards its goal, this is still an active line of research, and better techniques will help uncover the deeper educational gems hidden in how learners work through assignments.

Case study 3: MOOCs and multifaceted dropout factors. Big data inspires us to ask questions that we could not ask with previous types of educational data. Among these questions is whether we can predict learners' persistence in a course and understand the challenges they encounter, given data from their interactions with the system. In earlier learning environments, it was much easier to acquire data about a learner's skill through assessment tasks than it was to learn about the learner's motivation, volition, or other latent factors that affect persistence similarly. Newer online platforms record new types of interactions that make assessment of such latent factors more feasible. For instance, passive forum participation is a potential signal of motivation for learners who did not participate actively in the forum. Total time of a learner on the course site might be a signal of time availability.

This case study describes our attempt to leverage the richer types and scale of data to predict who is going to drop out from a MOOC, and whether they are going to drop out due to difficulty, lack of motivation, or lack of time. To predict who will drop out, we developed an algorithm that uses features extracted from learners' interactions with the videos, assignments, and forums in multiple MOOCs (Halawa, Greene, & Mitchell, 2014). Our model uses four features we found highly correlated with dropout: the amount of time taken to complete the first week's videos, assignment scores, and the fraction of videos and assignments skipped. The model predicted dropouts with a recall of 93% and false positive rate of 25%.

We developed an instrument and collected data to predict the reason(s) that learners drop out. We emailed a diagnostic survey to 9,435 learners who were red–flagged by our dropout predictor in a course. The survey was sent out via email in the middle of the third week of the course, and 808 recipients responded to the survey (a typical survey response rate in a MOOC). Constructing our diagnostic models based on the optional survey introduced a selection bias, whose consequences on the suitability of the designed interventions to non–respondents are the subject of future research. In the survey, learners were asked to report on various persistence factors, including their commitment level (the extent to which learners believed they committed a sufficient portion of their free time to the achievement of their course goals), and perceived difficulty (how difficult they found the course materials, including assessment tasks). Learners were also asked to report on the average amount of weekly free time they had. We used each learner's responses to compute three binary variables indicative of potential interventions:

1. Dropped out due to procrastination (which results from a lack of volition)
2. Dropped out due to difficulty
3. Dropped out due to lack of time

Next, we used learner interaction data to compute scores for various activity features describing the learner's pace, learning session times, and interactions with the lecture videos, assignments, and forums as shown in Table 1. We selected the features that we believe would correlate with particular reasons for dropout (or lack thereof). For instance, joining a study group may be predictive of the learner's intention to persist in the course for a long period. Giving up on problems after a first incorrect attempt might indicate a lack of motivation or grit.

Surprisingly, many reasonable statistics on learner trajectories are not particularly useful for predicting what expert teachers say is the correct way forward.

Table 1
Candidate Features Used to Predict Reasons for Dropout

Video interactions	• Fraction of visited video duration seeked back • Fraction of visited video duration seeked forward • Number of times the learner reviewed a previously visited video • Fraction of videos skipped • Fraction of videos viewed until the end
Assignment interactions	• Fraction of course quizzes attempted • Fraction of problems answered incorrectly on first attempt that were reattempted • Time between attempts
Forum interactions	• Number of forum posts • Number of comments to posts by other learners • Number of threads read • Did the learner post a self-introduction to the forum? • Did the learner create or join a study group?
Pace	• What fraction of released videos had the learner visited by various time points in the course?
Learning sessions	• How many times per week did the learner visit the course? • How long is the learner's average session?

We trained three logistic regression models, one for predicting each of the three dropout factors, which meant that a learner could be red–flagged for multiple dropout reasons. Accuracy was measured for each risk factor individually via recall – the fraction of learners who self–reported the risk factor that was red–flagged by the prediction model – and false positive rate (fpr) – the fraction of learners who were self–reportedly unaffected by the risk factor but red–flagged by the predictor (see Figure 2).

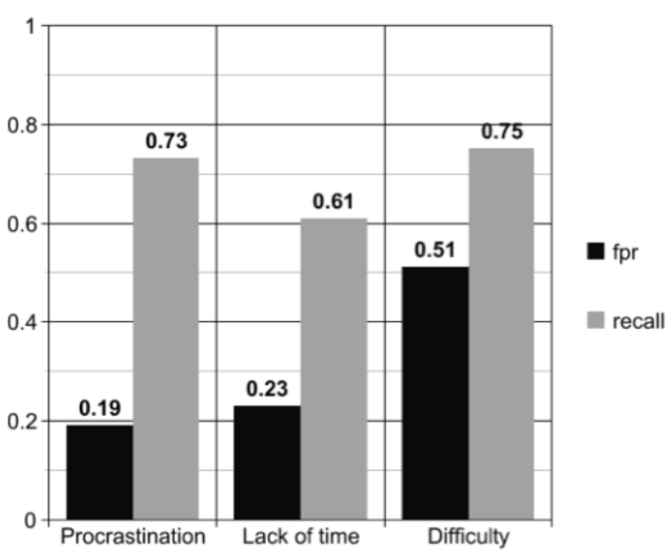


Figure 2. Prediction accuracy for our dropout diagnostic models

The procrastination detection model was able to predict procrastination with a false positive rate (fpr) of 0.19 at a recall of 0.73. The key contributing features of the model were interactions with the forum and assignments. We generally observed that learners with lower motivation or volition spent all of their time on the course in activities that yield direct personal rewards, such as viewing videos, taking assignments, reading the forum, or posting questions to the forum. Activities such as joining a study group, socializing on the forum, and commenting on other people's posts originated mainly from learners who self–reported higher levels of volition. We were also able to predict learners who reported time constraints with almost the same fpr but a lower recall. Contributing features included the patterns of spending time on the course, and it was observed that learners who report less free time tend to have shorter learning sessions. Predicting reports of perceived difficulty was less accurate due to the weakness of correlation between reported difficulty and our features including assignment scores. Improving this prediction is a subject of our future research.

This case study exemplifies two facets of data–enriched assessment, namely its multifaceted and feedback–oriented nature. In this study, we focused on specific facets of learners' contexts that are critical for their success in the learning environment: procrastination behavior, time constraints, and perceived difficulty. Moreover, this work will be extended to provide targeted feedback about these non–cognitive factors to at–risk learners. Potentially, such modeling capability allows us to assess these persistence factors and design more effective interventions that address the restraining and promoting forces relevant to each individual learner.

General Discussion

The preceding case studies illustrate how big data can enrich assessment by directly supporting learning as it assesses multiple facets of learning such as competencies and persistence. We argue that this is for three reasons. First, the next generation of online learning environments allows us to collect data continuously and at large scale. In turn, large–scale data collection allows researchers to more effectively use modern statistical and machine learning tools to identify and refine complex patterns of performance. For example, the work on programming trajectories described above illustrates that massive amounts of time–series data on learner programming problems can be used to predict later success and potentially to provide just–in–time hints.

Big data inspires us to ask questions that we could not ask with previous types of educational data.

Online learning environments also allow educators to record multifaceted measurements of skills and tendencies that normally evade traditional assessment tasks. The work on identifying dropout factors in MOOCs illustrates this point. Halawa and colleagues (2014) initially measured motivational variables using surveys, which are a familiar assessment instrument for academic motivation researchers. But they were then able to predict survey responses using data on forum engagement, pace, and other aspects of course interaction. In a traditional educational setting, these or analogous behavioral variables would be largely unmeasured. In addition, the continued development of educational games, complex simulations, and VR environments makes us confident that future educators will have a much more multifaceted set of data than ever before (Bailenson et al., 2008; Schwartz & Arena, 2013).

Third, and perhaps most crucially for learning, online learning environments are capable of delivering personalized feedback at the right moment. The Open Learning Initiative demonstrates this advantage by harnessing decades of research into cognitive skill development in order to model learner knowledge and provide more appropriate instruction in real time. Meta–analyses of what works in improving learning have placed appropriate feedback at or near the top of the list (Hattie, 2013), and researchers have argued that effective feedback is also the primary source of the oft–quoted "two–sigma" positive effects of tutoring (Bloom, 1984). Big data allows educators to build and refine model–driven feedback systems that can match and surpass human tutors (Corbett, 2001).

Finally, all of the examples in this article illustrate that big data can benefit multiple stakeholders in the learning ecosystem. As a more formative enterprise, data–enhanced assessment can benefit learners themselves, but it can also provide feedback to instructors to guide their attention and teaching strategies. The benefits of data–enriched assessment are available not only to instructors teaching in purely online environments but also to instructors teaching in hybrid (a blend of online and face to face instruction) or traditional classrooms. In hybrid environments, the data collected from the students in a class provide information to the instructor to make immediate adjustments to classroom teaching. Even instructors who are teaching in traditional classrooms without any technology will benefit from the insights about how students learn a subject that are developed from the big data collected in online learning environments. Big data have also clearly informed researchers to develop better learner models and experiment with just–in–time interventions. And Macfadyen, Dawson, Pardo, and Gaševic (2014) show that big data can inform questions about equitable and effective learning at a policy level.

Conclusion

We have been quite positive about the promise of data–enriched assessment, and so it seems reasonable to end with a note of caution. There is a difference between how we use assessment tasks and what they are intended to measure, and the history of psychometrics is littered with incorrectly interpreted test results. How will big data affect the interpretation and validity judgments of the next generation of assessment tasks? It may be helpful to look to the misapplication of current generation assessment tasks for lessons. Assessment experts generally agree that since the start of No Child Left Behind, data from high–stakes tests in K–12 settings have been used to make inaccurate judgments about the performance of teachers, schools, districts, and states in an attempt to establish benchmarks for accountability and quality improvement (Baker et al., 2010). According to a recent review, ten years of test–based accountability policies has shown little to no effects on student performance (National Research Council, 2011).

In earlier learning environments, it was much easier to acquire data about a learner's skill through assessment tasks than it was to learn about the learner's motivation, volition, or other latent factors that affect persistence similarly. Newer online platforms record new types of interactions that make assessment of such latent factors more feasible.

Exploring the network of causes for the misuse of standardized test data is beyond the scope of this paper, but there are two substantial causes worth noting that are deeply related to the tests themselves. The first is simply that our ambitions to capture learning have often outpaced our abilities to design effective assessment tasks – learning is a multifaceted construct that is difficult to measure. The second reason is that it is also difficult to appropriately aggregate, report, and act upon test data (National Research Council, 2011).

We have argued that a data–enriched assessment process can potentially measure multiple facets of learning, as well as learning processes, more effectively than previous assessment approaches. However, our case studies also show that these assessment tasks depend on broad and deep learner data that may not always be available. The hype around online assessment, and the excitement over measuring novel motivational and other non–cognitive competencies, may continue to fuel ambitions that outstrip our capabilities. Moreover, data–enriched assessment methods can be far more complex and opaque than traditional methods, and their results can be difficult to interpret without expert assistance (Siemens & Long, 2011).

The availability of big data allows assessment methods to continually measure and support a broader range of learning outcomes while simultaneously providing feedback throughout the learning process. This is creating more of a need to provide thoughtful and actionable explanations of assessment results for all of the stakeholders involved, including teachers and learners.

AUTHOR'S NOTE

This work is a collaborative work by the researchers in the Stanford University Lytics Lab (http://lytics.stanford.edu). Each listed author is an equal contributor to the work.

References

Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences, 17*(1), 102–141.

Baker, E. L., Barton, P. E., Darling–Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., & Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers. EPI briefing paper# 278. *Economic Policy Institute.* Retrieved from http://eric.ed.gov/?id=ED516803

Baker, R. S., & Corbett, A. T. (2014). Assessment of robust learning with educational data mining. *Research & Practice in Assessment, 9*(2), 38-50.

Black, P., & Williams, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one–to–one tutoring. *Educational Researcher, 13*(6), 4–16.

Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self–affirmation and social psychological intervention. *Annual Review of Psychology, 65*, 333–371.

Corbett, A. (2001). Cognitive computer tutors: Solving the two–sigma problem. In M. Bauer, P. J. Gmytrasiewicz, & J. Vassileva, (Eds.), *User modeling* (pp. 137–147). Heidelberg, Germany: Springer.

De Laat, M., & Prinsen, F. R. (2014). Social learning analytics for higher education. *Research & Practice in Assessment, 9*(2) 51-60.

Denley, T. (2014). How predictive analytics and choice architecture can improve student success. *Research & Practice in Assessment, 9*(2), 61-69.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Ballantine Books.

Gikandia, J. W., Morrowa, D., & Davisa, N. E. (2011). Online formative assessment in higher education: A review of the literature. *Computers & Education, 57*(4), 2333–2351.

Halawa, S., Greene, D., & Mitchell, J. (2014). Dropout prediction in MOOCs using learner activity features. *Proceedings of the European MOOC Summit*. Lausanne, Switzerland.

Hattie, J. (2013). *Visible learning: A synthesis of over 800 meta–analyses relating to achievement*. New York, NY: Routledge.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. *Proceedings of the Third International Conference on Learning Analytics and Knowledge, ACM* (pp. 170–179).

Macfadyen, L. P., Dawson, S., Pardo, A., & Gaševic, D. (2014). The learning analytics imperative and the sociotechnical challenge: Policy for complex systems. *Research & Practice in Assessment, 9*(2), 17–28.

National Research Council. (2011). *Incentives and test–based accountability in public education*. Committee on Incentives and Test–Based Accountability in Public Education, M. Hout & S. W. Elliott, (Eds.). Board on Testing and Assessment, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academies Press.

Pintrich, P. R. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology, 95*(4), 667–686.

Roediger, H. L., & Karpicke, J. D. (2006). Test–enhanced learning: Taking memory tests improves long–term retention. *Psychological Science, 17*(3), 249–255.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. (Eds.). (2011). *Handbook of educational data mining*. Boca Raton, FL: CRC Press.

Schwartz, D. L., & Arena, D. (2013). *Measuring what matters most: Choice–based assessments for the digital age.* Cambridge, MA: The MIT Press.

Shavelson, R. J. (2007). *A brief history of student learning assessment: How we got to where we are and where to go next*. Washington, DC: Association of American Colleges and Universities.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

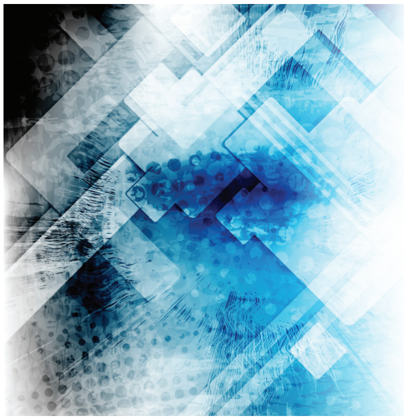
Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46*(5), 30–32.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613–629.

Walton, G. M., & Cohen, G. L. (2011). A brief social–belonging intervention improves academic and health outcomes of minority students. *Science, 331*(6023), 1447–1451.

Ware, C. (2013). *Information visualization: Perception for design* (3rd ed.). Waltham, MA: Elsevier.

Zimmerman, B. J. (1990). Self–regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 1–25.



Abstract

In the new era of big educational data, learning analytics (LA) offer the possibility of implementing real–time assessment and feedback systems and processes at scale that are focused on improvement of learning, development of self–regulated learning skills, and student success. However, to realize this promise, the necessary shifts in the culture, technological infrastructure, and teaching practices of higher education, from assessment–for–accountability to assessment–for–learning, cannot be achieved through piecemeal implementation of new tools. We propose here that the challenge of successful institutional change for learning analytics implementation is a wicked problem that calls for new adaptive forms of leadership, collaboration, policy development and strategic planning. Higher education institutions are best viewed as complex systems underpinned by policy, and we introduce two policy and planning frameworks developed for complex systems that may offer institutional teams practical guidance in their project of optimizing their educational systems with learning analytics.

AUTHORS

Leah P. Macfadyen
The University of British Columbia

Shane Dawson
University of South Australia

Abelardo Pardo
The University of Sydney

Dragan Gašević
Athabasca University

Embracing Big Data in Complex Educational Systems: The Learning Analytics Imperative and the Policy Challenge

In education, we are awash in data about our learners and educators, our technologies and activities, achievements and performance. To date these data have rarely been mined intelligently with the goal of improving learning and informing teaching practice, although evidence from other sectors such as marketing, sports, retail, health and technology suggests that the effective use of big data can offer the education sector the potential to enhance its systems and outcomes (Manyika et al., 2011). Norris and Baer (2013) have noted that, "Data expands the capacity and ability of organizations to make sense of complex environments" (p. 13). In this context, learning analytics (LA) offers the capacity to investigate the rising tide of learner data with the goal of understanding the activities and behaviors associated with effective learning, and to leverage this knowledge in optimizing our educational systems (Bienkowski, Feng, & Means, 2012; Campbell, DeBlois, & Oblinger, 2007). Indeed, in a world of larger and larger data sets, increasing populations of increasingly diverse learners, constrained education budgets and greater focus on quality and accountability (Macfadyen & Dawson, 2012), some argue that using analytics to optimize learning environments is no longer an option but an imperative. The value of such analytics is highlighted by the authors of the McKinsey Global Institute (Manyika et al., 2011) noting that, "In a big data world, a competitor that fails to sufficiently develop its capabilities will be left behind…Early movers that secure access to the data necessary to create value are likely to reap the most benefit" (p. 6). Education can no longer afford not to use learning analytics. As Slade and Prinsloo (2013) maintain, "Ignoring information that might actively help to pursue an institution's goals seems shortsighted to the extreme" (p. 1521).

CORRESPONDENCE

Email
leah.macfadyen@ubc.ca

In this article we consider ways in which learning analytics can support and contribute to the development of new approaches to the assessment of learning, and the degree to which new adaptive policy and planning approaches will be needed to bring about the kind of institutional change such proposals demand. We emphasize that successful institutional adoption demands comprehensive development and implementation of policies to address LA challenges of learning design, leadership, institutional culture, data access and security, data privacy and ethical dilemmas, technology infrastructure, and a demonstrable gap in institutional LA skills and capacity (Siemens, Dawson, & Lynch, 2013). Moreover, we take the position that educational institutions are complex adaptive systems (Gupta & Anish, 2009; MacLennan, 2007; Mitleton–Kelly, 2003), and therefore that simplistic approaches to policy development are doomed to fail. Instead, we will introduce strategy and policy frameworks and approaches developed for complex systems, including frameworks that offer the potential to identify points of intervention (Corvalán, Kjellström, & Smith, 1999), with the goal of offering educational institutions practical guidance.

Assessment Practices: A Wicked Problem in a Complex System

Indeed, in a world of larger and larger data sets, increasing populations of increasingly diverse learners, constrained education budgets and greater focus on quality and accountability, some argue that using analytics to optimize learning environments is no longer an option but an imperative.

There is no better exemplar in higher education than assessment to demonstrate how institutional policy can impact practice both positively and negatively. The practice of assessment has for some time been mired in debate over its role as either a measure of accountability or a process for learning improvement. While the majority of education practitioners lean towards assessment as a process for improving student learning, assessment nonetheless remains firmly positioned as an important tool for determining accountability and demonstrating quality. As McDonnell (1994) previously argued, assessment policies function as a mechanism to provide government with a high level of influence over classroom practice. In essence, assessment acts as a powerful tool to manage aspects of learning and teaching. It is not surprising, then, that assessment policy has numerous invested stakeholders – learners, educators, administrators and government – all vying for a larger stake in the game. The diversity of stakeholders, priorities, outcomes and needs make any substantial change to assessment policy and practice a considerable challenge to say the least.

Assessment practice will continue to be intricately intertwined both with learning and with program accreditation and accountability measures. Such interconnectedness in educational systems means that narrow efforts to implement changes in policy and practice in one area (for example, by introducing new approaches to tracking and measuring learning) may have unanticipated consequences elsewhere in the system. For example, the US education policy No Child Left Behind drastically reshaped not only the testing processes employed to identify poor literacy and numeracy standards, but also affected what was taught and how it was taught. Jacob (2005) documented the unintentional outcomes of this new accountability policy classroom practice, noting, for example that such high–stakes testing encouraged teachers to steer low–performing students away from subjects that were included in the accountability program. While the ethos of the policy had some merit in attempting to address declining numeracy and literacy skills in the US, the associated incentives and measures resulted in crossed performance indicators. Dworkin (2005) also expands on this point, noting that teacher promotion standards were linked to class performance in the high stakes tests. This practice essentially encouraged teachers to narrow the curriculum and teach to the test, beautifully illustrating Goodhart's Law, which states that when a measure becomes a target it ceases to be a useful measure (Elton, 2004).

In the complex systems of higher education, current performance assessment and accountability policies may be the forces driving (Corvalán et al., 1999) the continued focus on high–stakes snapshot testing as a means of producing comparative institutional data, in spite of the well–articulated weakness of such an approach for understanding student learning. The continuing primary use of grades in determining entry to university, the Australian Government's National Assessment Plan for Literacy and Numeracy (NAPLAN)[1] measures, the OECD's Programme for International Student Assessment (PISA)[2] and similar programs, demonstrate that there is much invested in the retention of these measures for benchmarking individuals, schools, districts, states and countries. Wall, Hursh and Rodgers (2014) have

[1] http://education.qld.gov.au/naplan/

[2] http:/www.oecd.org/pisa/

argued, on the other hand, that the perception that students, parents and educational leaders can only obtain useful comparative information about learning from systematized assessment is a false one. Instead, alternate complementary assessment practices – practices that make better use of the rich array of educational data now available – may well offer more effective approaches to improving learning, especially processes that reveal development of student understanding over time (Wiliam, 2010).

In his criticism of assessment practices, Angelo (1999) suggested that as educators we must emphasize assessment as a means for improving student learning rather than a mechanistic, technical process used to monitor performance. He argued that assessing for learning necessitates a focus on developing practices that help the educator and learner gather evidence of learning progress, rather than on identifying the students that produce the "right" or "wrong" answers. The importance of developing better formative or embedded assessment models has also been reiterated by the OECD Innovative learning environments project (Dumont, Istance, & Benavides, 2010) and educational researchers have similarly illuminated that regular feedback at the process level is more effective for enhancing deeper learning (for review, see Hattie & Timperley, 2007).

Despite the widespread recognition of the need for a more effective assessment paradigm, implementation is a challenge, and calls for development of new policies and implementation strategies directed at improving accountability for learning though practices driven by learning. Differentiating assessment–for–learning from assessment–for–accountability within the educational system forms part of the wicked problem of institutional change in higher education that we seek to explore here. As with all complex systems, even a subtle change may be perceived as difficult, and be resisted (Head & Alford, 2013). For example, under normal classroom circumstances the use of assessment at the process level for improving learning requires substantial and sustained engagement between the educator and students and can be an extremely time intensive process. Implementing such time intensive assessment models for large (and growing) university classes is not feasible, and typically scalable models of assessment such as multiple choice exams are implemented instead. It is unrealistic to consider that educators will adopt time–consuming longitudinal and personalized assessment models given the massive increase in resources and workload that would be required.

Learning Analytics and Assessment–for–Learning

A wide range of authors in this special issue illustrate ways in which learning analytics – which comes with its own set of implementation challenges and hurdles – has the potential to provide learners with sustained, substantial and timely feedback to aid understanding and improve student learning skills, while circumventing the challenge of educator workload. We also offer a discussion of how learning analytics may support development of self–regulated learning in Box 1, inset. Analytics can add distinct value to teaching and learning practice by providing greater insight into the student learning process to identify the impact of curriculum and learning strategies, while at the same time facilitating individual learner progress. Nor does the adoption of learning analytics preclude traditional or alternate assessment practices that may be required by accreditation and accountability policies. While current assessment policy may be driven by conflicting intentions – accountability and quality assurance requirements versus promotion of student learning – learning analytics can meet both. More simply put, LA addresses the need for quality assurance and learning improvement.

There is no better exemplar in higher education than assessment to demonstrate how institutional policy can impact practice both positively and negatively. The practice of assessment has for some time been mired in debate over its role as either a measure of accountability or a process for learning improvement.

Technological Components of the Educational System and Support of Learning Analytics

The LA–supported approaches to assessment of learning envisioned in this article – indeed, in this entire edition – assumes a technological layer that is capable of capturing, storing, managing, visualizing and processing big educational data – the millions of events occurring in diverse learning scenarios and platforms. Transformation of assessment practices to embrace and integrate learning analytics tools and strategies in support of teaching and learning therefore demands effective institutional technology infrastructures. The production of data in every technology–mediated interaction occurring in a learning environment, the need for more effective provision of feedback, and the need for more

Box 1
Learning Analytics for Assessing Student Learning

Differentiating assessment–for–learning from assessment–for–accountability within the educational system forms part of the wicked problem of institutional change in higher education that we seek to explore here.

Provision (to learners and educators) of automated analytics that provide feedback on learner study behaviors, progress and outcomes will not only enhance academic performance but also develop student self–regulated learning (SRL) skills, and SRL proficiency has been demonstrated to be a significant predictor of academic success (e.g., Butler & Winne, 1995; Pintrich, 1999; Zimmerman, 2002). Student motivation and capacity to undertake accurate self–monitoring had a direct impact on the level and quality of their study and therefore, their overall learning progression and academic achievement (Dunlosky & Thiede, 1998). Conversely, poor performers are poor at evaluating their own ability or judging their own learning skills (Kruger & Dunning, 1999). For these reasons, it is argued that a core goal of any effective pedagogical strategy must include the development of student meta–cognitive skills or judgment of (own) learning (JOL). Feedback on assessment is one key approach that is often adopted to assist students in developing meta–cognitive skills, but because provision of relevant feedback can be labor–intensive, it is often delayed and provided at a time when it is no longer useful to the student to aid their learning.

Recent research posits that SRL is a process of temporal events that evolve during learning (Azevedo & Aleven, 2013). This research, alongside recent developments in learning analytics, data mining and machine learning is providing new methods for developing novel insights into student learning processes. Historically, assessment and development of student SRL has made use of tasks associated with JOL which generally involve asking a student to assess how effectively they have understood a particular concept (Dunlosky & Lipko, 2007). This self–reported rating is then correlated against their overall test performance to gain insight into the student's meta–cognitive proficiency. While JOL has commonly relied on self–report methodologies such as think aloud protocols and surveys, these have inherent limitations such as poor recall, and biased responses (Richardson, 2004).

New options for assessing student learning behaviors are emerging as a result of advances in learning analytics and natural language processing (NLP), and alternate models have sought to capture actual learner behavior (in lieu of self–reports) from interactions with technology–based learning activities. For example, oft–cited SRL researcher Phil Winne has previously reported that student online interaction data can provide significant indicators of SRL proficiency (e.g., Winne, 2010; Zhou & Winne, 2012). Winne has developed the software application nStudy as a web tool that can collect very fine grained, time stamped data about individual learner interactions with online study materials. The trace data is then used to provide insight and feedback into the learner's cognitive choices as they interact with the online information. Essentially, the tool makes data for reflection available to both the individual learner and the educator.

comprehensive formative and summative assessment translates into a rich set of requirements of the current technological infrastructures. Although learning management systems (LMSs) still host a large percentage of technology–mediated educational activities, educational institutions are recognizing the need to re–assess the concept of teaching and learning space to encompass both physical and virtual locations, and adapt learning experiences to this new context (Thomas, 2010). Thus, together with the need for cultural change and a focus on pedagogical relevance, an additional sociotechnical factor critical to the adoption of learning analytics is technology itself (Box 2).

The evolution of technology in recent years offers an unprecedented capacity to store large data sets, and applications using big data are well established in contexts such as business intelligence, marketing and scientific research (Dillon, Wu, & Chang, 2010). Education faces a particular challenge that derives from the rich variety of technological affordances emerging in learning environments. From an LMS–centric approach consolidated in the early 2000s, we are now entering an era in which learning may occur anywhere, at any time, with multiple devices, over a highly heterogeneous collection of resources, and through multiple types of interactions. In this new scenario, learning analytics tools need to comply with requirements in the following areas:

1. Diverse and flexible data collection schemes: Tools need to adapt to increasing data sources, distributed in location, different in scope, and hosted in any platform.

2. Simple connection with institutional objectives at different levels: information needs to be understood by stakeholders with no extra effort. Upper management needs insight connected with different organizational aspects than an educator. User–guided design is of the utmost importance in this area.

3. Simple deployment of effective interventions, and an integrated and sustained overall refinement procedure allowing reflection.

From the technological point of view, learning analytics is an emerging discipline (Siemens, 2013) and its connection with assessment remains largely unexplored (Ellis, 2013). This situation is even more extreme when considering the assessment of competences and learning dispositions (Buckingham Shum, 2012). Educational institutions need technological

Box 2
Sociotechnical Infrastructure Needs for Effective Learning Analytics

Several initiatives are already tackling the problem of flexible data collection schemes. For example the ADL Experience API[3] released in 2013 has been proposed as a solution that can promote interoperability between data collected in different environments and platforms. The interface offers the possibility of capturing a wide variety of events in experiences with heterogeneous scenarios (Glahn, 2013). Similarly, the IMS Global Consortium has proposed that the Learning Measurement Framework IMS Caliper[4] – containing descriptions to represent metrics, sensor API and learning events – will facilitate the representation and processing of big data in the learning field. In parallel, the concept of a *Learning Record Store* (LRS) has been proposed as a framework for storing and manipulating data from distributed events in a learning environment, encoding not only interaction among stakeholders, but among resources. This information is then made available through a service–based interface to other systems within an institution (or across multiple institutions) for further analysis and processing.

Numerous attempts have been made to meet diverse stakeholder reporting and data access needs by production of so–called dashboards that show a canvas of multiple visualizations. Common limitation of these graphical representations, however, are their actual utility and usability (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013). Adapting presentation of information to user context, needs and interests is another important factor that must be taken into account if we wish to facilitate the uptake of learning analytics solutions.

The third requirement for technology supporting learning analytics is that it can facilitate the deployment of so–called interventions, where intervention may mean any change or personalization introduced in the environment to support student success, and its relevance with respect to the context. This context may range from generic institutional policies, to pedagogical strategy in a course. Interventions at the level of institution have been already studied and deployed to address retention, attrition or graduation rate problems (Ferguson, 2012; Fritz, 2011; Tanes, Arnold, King, & Remnet, 2011). More comprehensive frameworks that widen the scope of interventions and adopt a more formal approach have been recently proposed, but much research is still needed in this area (Wise, 2014).

It is unrealistic to consider that educators will adopt time–consuming longitudinal and personalized assessment models given the massive increase in resources and workload that would be required.

[3] http://www.adlnet.gov/tla

[4] http:/www.imsglobal.org/IMSLearningAnalyticsWP.pdf

solutions that are deployed in a context of continuous change, with an increasing variety of data sources, that convey the advantages in a simple way to stakeholders, and allow a connection with the underpinning pedagogical strategies.

In turn, these technological requirements point to a number of other critical contextual factors that must form part of any meaningful policy and planning framework for employing learning analytics in service of improved assessment. Foremost among these is the question of access to data, which needs must be widespread and open. Careful policy development is also necessary to ensure that assessment and analytics plans reflect the institution's vision for teaching and strategic needs (and are not simply being embraced in a panic to be seen to be doing something with data), and that LA tools and approaches are embraced as a means of engaging stakeholders in discussion and facilitating change rather than as tools for measuring performance or the status quo.

The Challenge: Bringing about Institutional Change in Complex Systems

While the vision of improving student learning and assessment through implementation of effective learning analytics tools and approaches holds promise, the real challenges of implementation are significant. In this article we have identified only two of the several critical and interconnected socio–technical domains that need to be addressed by comprehensive institutional policy and strategic planning to introduce such attractive new systems: the challenge of influencing stakeholder understanding of assessment in education, and the challenge of developing the necessary institutional technological infrastructure to support the undertaking. Meanwhile, of course, any such changes must coexist with the institution's business as usual obligations (Head & Alford, 2013).

It may not be surprising, then, that globally, education lags behind all other sectors in harnessing the power of analytics. A preliminary analysis indicates that educational institutions simply lack the practical, technical and financial capacity to effectively gather, manage and mine big data (Manyika et al., 2011). As Bichsel (2012) notes, much concern revolves around "the perceived need for expensive tools or data collection methods" (p. 3). Certainly, evidence suggests that data access and management are proving to be significant hurdles for many institutions. The first survey of analytics implementation in US higher education in 2005 found that of 380 institutions, 70% were at Stage 1 of a five–stage implementation process: "Extraction and reporting of transaction–level data" (Goldstein & Katz, 2005). Four years later, a study of 305 US institutions found that 58% continued to wrangle data in Stage 1, while only 20% reported progress to Stage 2: "Analysis and monitoring of operational performance" (Yanosky, 2009). More recently, investigators have reported that while some 70% of surveyed institutions agreed that analytics is a major priority for their school, the majority of respondents suggested that data issues (quality, ownership, access, and standardization) were considerable barriers to analytics implementation, and as such most were yet to make progress beyond basic reporting (Bichsel, 2012; Norris & Baer, 2013).

Transformation of assessment practices to embrace and integrate learning analytics tools and strategies in support of teaching and learning therefore demands effective institutional technology infrastructures

To further unpack the complexities of analytics adoption a growing number of commentators are exploring the more nuanced sociotechnical factors that are the most likely barriers to institutional LA implementation. For instance, elements of institutional "culture, capacity and behavior" (Norris, Baer, Leonard, Pugliese, & Lefrere, 2008). There is recognition that even where technological competence and data exist, simple presentation of the facts (the potential power of analytics), no matter how accurate and authoritative, may not be enough to overcome institutional resistance (Macfadyen & Dawson, 2012; Young & Mendizabal, 2009).

Why Policy Matters for Learning Analytics

Higher education institutions are a superb example of complex adaptive systems (CASs) (Cilliers, 1998; Gupta & Anish, 2009; MacLennan, 2007; Mitleton–Kelly, 2003) and exist in a state that some have characterized as organized anarchy (Cohen & Marsh, 1986). Together with institutional history and differences in stakeholder perspectives (Kingdon, 1995; Sabatier, 2007), policies are the critical driving forces that underpin complex and systemic institutional problems (Corvalán et al., 1999) and that shape perceptions of the nature of the problem(s) and acceptable solutions. Below, we argue that it is therefore only through implementation of planning processes driven by new policies that institutional change can come about.

The challenge of bringing about institution–wide change in such complex and anarchic adaptive systems may rightly be characterized as a "wicked problem"– a problem that is "complex, unpredictable, open ended, or intractable" (Churchman, 1967; Head & Alford, 2013; Rittel & Webber, 1973). Like all complex systems, educational systems are very stable, and resistant to change. They are resilient in the face of perturbation, and exist far from equilibrium, requiring a constant input of energy to maintain system organization (see Capra, 1996). As a result, and in spite of being organizations whose business is research and education, simple provision of new information to leaders and stakeholders is typically insufficient to bring about systemic institutional change. One factor hindering institutional change for better use of analytics by educational institutions appears to be their "lack of data–driven mind–set and available data" (Manyika et al., 2011, p. 9). Interestingly, this observation is not new, and was reported with dismay in 1979 by McIntosh, in her discussion of the failure of institutional research to inform institutional change. Ferguson et al. (in press) reprise McIntosh's arguments in relation to learning analytics, suggesting that additional barriers to adoption include academics' unwillingness to act on findings from other disciplines; disagreement over the relative merits of qualitative vs. quantitative approaches to educational research; a tendency to base decisions on anecdote; the reality that researchers and decision makers speak different languages; lack of familiarity with statistical methods; a failure to effectively present and explain data to decision makers; and the reality that researchers tend to hedge and qualify conclusions. Norris and Baer (2013) meanwhile note that the analytics IQ of institutional leaders is typically not high, precluding effective planning. In other words, a range of political, social, cultural and technical norms shape educational systems and contribute to their stability and resistance to change.

…we are now entering an era in which learning may occur anywhere, at any time, with multiple devices, over a highly heterogeneous collection of resources, and through multiple types of interactions.

Elsewhere, we reported on a case study failure of learning analytics to inform institutional planning (Macfadyen & Dawson, 2012), and noted that the culture of educational institutions has historically valorized educator/faculty autonomy and resisted any administrative efforts perceived to interfere with teaching and learning practice. We proposed that in order to overcome institutional resistance to innovation and change driven by learning analytics, educational institutions urgently need to implement planning processes that create conditions that allow stakeholders across the institution to both think and feel positively about change – conditions that appeal to both the heart and the head.

Social marketing theorists (Kotler & Zaltman, 1971) and change management experts (Kavanagh & Ashkanasy, 2006; Kotter, 1996) similarly argue that social and cultural change (that is, change in habits, practices and behaviors) is not brought about by simply giving people large volumes of logical data (Kotter & Cohen, 2002). Social theorists have argued that since value perspectives ground the major social issues of modern life, scientific analyses and technical rationality are insufficient mechanisms for understanding and solving complex problems (Head & Alford, 2013; Rein, 1976; Schon & Rein, 1994). Instead, what is needed are comprehensive policy and planning frameworks to address not simply the perceived shortfalls in technological tools and data management, but the cultural and capacity gaps that are the true strategic issues (Norris & Baer, 2013).

Policy and Planning Approaches for Wicked Problems in Complex Systems

Policies are, simply, principles developed to guide subjective and/or objective decision making, with the goal of achieving rational and desirable outcomes. They are statements of intent that capture organizational goals, and are typically implemented via planned procedures or protocols. A large and established literature on policy development already exists in fields such as political science and business, from which have emerged a range of classical policy cycle tools and heuristics that have been highly influential (Nakamura, 1987). Contemporary critics from the planning and design fields argue, however, that these classic, top–down, expert–driven (and mostly corporate) policy and planning models are based on a poor and homogenous representation of social systems mismatched with our contemporary pluralistic societies, and that implementation of such simplistic policy and planning models undermines chances of success (for review, see Head & Alford, 2013). Importantly, they also insist that modern policy problems are not technical puzzles that can be solved through the application of scientific knowledge, but instead exist in continuous states of flux within dynamic systems and have communicative, political and institutional elements. Solutions to such ill–defined and multi–factorial challenges, they argue, will always be provisional, and must be negotiated between multiple stakeholders in situations of ambiguity, uncertainty and values disagreement (Rittel & Webber, 1973). A number of theorists have also emphasized that solutions to wicked problems – actually complex systems of inter–related problems – "can seldom be obtained by independently solving each of the problems of which it is composed . . . Efforts to deal separately with such aspects of urban life as transportation, health, crime, and education seem to aggravate the total situation" (Ackoff, 1974, p. 21).

Systems theory offers two key areas of insight that are significant for policy development for complex educational systems. First, systems theorists recognized that while systems – from a single atom to a universe – may appear to be wildly dissimilar, they are all governed by common patterns, behaviors and properties: their component parts are multiply interconnected by information flows, with identifiable and predictable feedbacks, inputs, outputs, controls and transformation processes; they are dynamic, differentiated and bounded; they are hierarchically organized and differentiated; and new properties can arise within them as a result of interactions between elements. Second, systems theory observes that systems tend to be stable, and that their interconnectedness facilitates resilience (for a review of systems theory, see Capra, 1996).

From the technological point of view, learning analytics is an emerging discipline and its connection with assessment remains largely unexplored.

These observations not only illuminate why piecemeal attempts to effect change in educational systems are typically ineffective, but also explains why no one–size–fits–all prescriptive approach to policy and strategy development for educational change is available or even possible. Usable policy frameworks will not be those which offer a to do list of, for example, steps in learning analytics implementation. Instead, successful frameworks will be those which guide leaders and participants in exploring and understanding the structures and many interrelationships within their own complex system, and identifying points where intervention in their own system will be necessary in order to bring about change.

Drawing on systems and complexity theory, a new generation of authors have begun to develop accounts of so–called adaptive approaches to policy and planning for complex systems which can allow institutions to respond flexibly to ever–changing social and institutional contexts and challenges (Berkhout, Leach, & Scoones, 2003; Haynes, 2003; Milliron, Malcolm, & Kil, 2014; Tiesman, van Buuren, & Gerrits, 2009; Young & Mendizabal, 2009). A full review of adaptive management strategies is beyond the scope of this paper, and has been comprehensively undertaken by Head and Alford (2013), who highlight the critical roles of cross–institutional collaboration, new forms of leadership (moving beyond the orthodox model of transformational leadership) and the development of enabling structures and processes (for example, budgeting and finance systems, organizational structure, human resources management, and approaches to performance measurement and program evaluation). We offer here two sample policy and planning models that may offer valuable practical guidance for collaborative teams and leaders in higher education seeking to bring about systemic institutional change to support learning analytics.

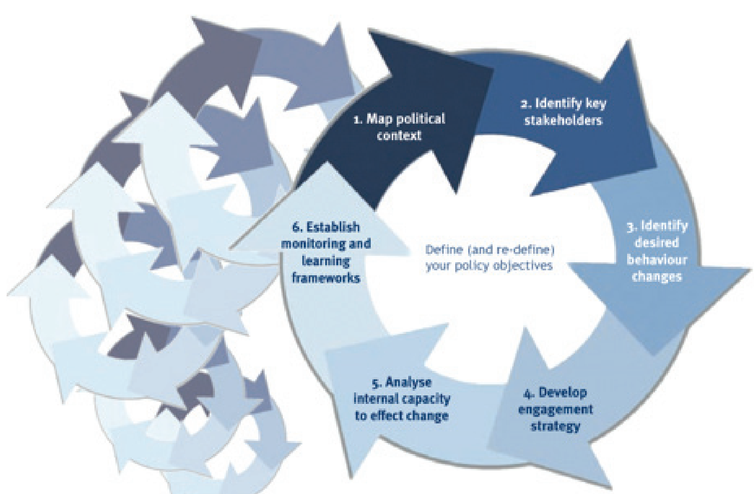


Figure 1. The RAPID Outcome Mapping Approach (ROMA)

First, and as we have proposed elsewhere (Ferguson et al., in press) we offer a modification of Young and Mendizabal's (2009) Rapid Outcome Mapping Approach (ROMA) model (Figure 1) as a policy and planning heuristic for learning analytics implementation. Originally developed to support policy and strategy processes in the complex field of international development, the seven–step ROMA model is focused on evidence–based policy change. It is designed to be used iteratively, and to allow refinement and adaptation of policy goals and the resulting strategic plans over time and as contexts change, emphasizing the provisional nature of any solutions arrived at. Importantly, the ROMA process begins with a systematic effort at mapping institutional context (for which these authors offer a range of tools and frameworks) – the people, political structures, policies, institutions and processes that may help or hinder change. This critical activity allows institutions to identify the key factors specific to their own context that may influence (positively or negatively) the implementation process, and therefore also has the potential to illuminate points of intervention and shape strategic planning.

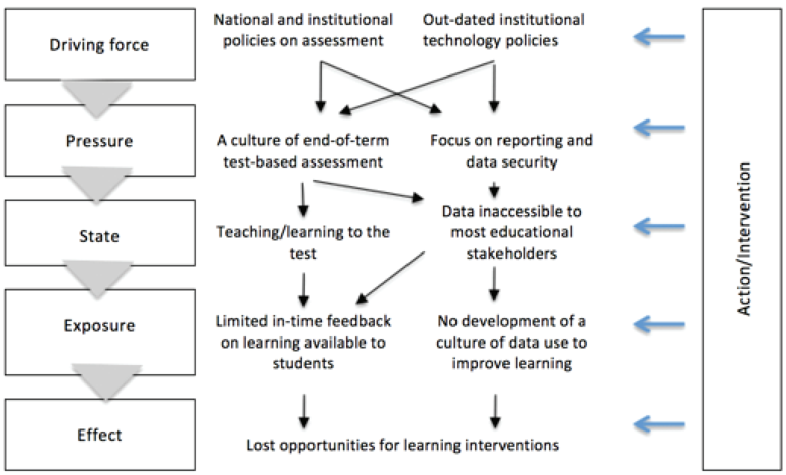


Figure 2. Cause–effect (DPSEEA) framework for institutional assessment and technology policies (modified from Corvalan et al., 1999).

Second, Corvalán et al.'s (1999) "cause–effect framework" (or DPSEEA framework) usefully assists in identifying the multiple linkages that may exist between the driving forces underpinning complex systems, illuminating the multiple points in a complex system of relationships where action may be needed to effect change. Such a framework can, they suggest, "be used to weigh alternatives and to design step–by–step programs for dealing with a particular…problem" (p. 659). Figure 2 offers a preliminary modification of this framework to represent institutional effects of, for example, technology and assessment policies, and may be a useful context mapping tool in the ROMA process.

It may not be surprising, then, that globally, education lags behind all other sectors in harnessing the power of analytics. A preliminary analysis indicates that educational institutions simply lack the practical, technical and financial capacity to effectively gather, manage and mine big data.

Use of these models for institutional LA policy development is only in the very early stages, although we have explored elsewhere (Ferguson et al., in press) the ways in which a small number of apparently successful institutional LA policy and planning processes have pursued change management approaches that map well to such frameworks. In future work, we hope to be able to present more robust and critical review of real–time application of these frameworks in institutional planning, and their possible effectiveness or limitations.

In the meantime, readers may review both frameworks and immediately dispute the stages, levels, linkages, effects or impacts in relation to their own institutional context. But this is, of course, the very point of such adaptive models, which can and should be disputed, negotiated and modified as needed for local institutional contexts, to guide relevant local action. To paraphrase Head and Alford (2013), when it comes to wicked problems in complex systems, there is no one–size–fits–all policy solution, and there is no plan that is not provisional.

Rather, the more important role of such frameworks is to continuously remind us of the need for a holistic understanding of institutional context if the goal is institutional change, including external and internal influences, political and cultural context, the evidence itself, and the links:

"All of the other actors and mechanisms that affect how the evidence gets into the policy process" (Young & Mendizabal, 2009). They can assist in identifying points of intervention (Corvalán et al., 1999) in the complex adaptive system that is education, to offer leaders and practitioners additional insight and tools in their project of optimizing the system with learning analytics.

References

Ackoff, R. L. (1974). *Redesigning the future*. New York, NY: Wiley.

Angelo, T. A. (1999). Doing assessment as if learning matters most. *AAHE Bulletin, 51*(9), 3–6.

Azevedo, R., & Aleven, V. (Eds.). (2013). *International handbook of metacognition and learning technologies.* Amsterdam: Springer.

Berkhout, F., Leach, M., & Scoones, I. (Eds.). (2003). *Negotiating environmental change: New perspectives from social science*. Cheltenham, UK: Edward Elgar.

Bichsel, J. (2012). *Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations* (Research Report). Louisville, CO: EDUCAUSE Center for Applied Research. http://net.educause.edu/ir/library/pdf/ERS1207/ers1207.pdf

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief.* Washington, DC: U.S. Department of Education Office of Educational Technology. http://www.ed.gov/edblogs/technology/files/2012/03/edm–la–brief.pdf

Buckingham Shum, S. (2012). Learning dispositions and transferable competencies: Pedagogy, modelling and learning analytics. In S. Buckingham Shum, D. Gašević, & R. Ferguson (Eds.), *International conference on learning analytics and knowledge* (pp. 92–101). New York, NY: ACM Press.

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. G. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review, 42*(4), 42–57.

Capra, F. (1996). *The web of life*. New York, NY: Doubleday.

Churchman, C. W. (1967). Free for all. *Management Science, 14*, B141–B142.

Cilliers, P. (1998). *Complexity and postmodernism: Understanding complex systems*. London, UK: Routledge.

Cohen, M. D., & Marsh, J. G. (1986). *Leadership and ambiguity: The American college president*. New York, NY: McGraw Hill.

Corvalán, C. F., Kjellström, T., & Smith, K. R. (1999). Health, environment and sustainable development: Identifying links and indicators to promote action. *Epidemiology, 10*(5), 656–660.

Dillon, T., Wu, C., & Chang, E. (2010). Cloud computing: Issues and challenges. In *IEEE International Conference on Advanced Information Networking and Applications* (pp. 27–33). New York, NY: IEEE Press.

Dumont, H., Istance, D., & Benavides, F. (Eds.). (2010). The nature of learning: Using research to inspire practice. *Educational Research and Innovation series*, OECD Publishing. http://www.oecd.org/edu/ceri/thenatureoflearningusingresearchtoinspirepractice.htm

Dunlosky, J., & Lipko, A. R. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science, 16*(4), 228–232.

Dunlosky, J., & Thiede, K. W. (1998). What makes people study more? An evaluation of factors that affect self–paced study. *Acta Psychologica, 98*(1), 37–56.

Dworkin, G. (2005). The No Child Left Behind act: Accountability, high–stakes testing and roles for sociologists. *Sociology of Education, 78*(2), 170–174.

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.

Elton, L. (2004). Goodhart's Law and performance indicators in higher education. *Evaluation & Research in Education, 18*(1–2), 120–128.

Ferguson, R. (2012). Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6), 304–317.

Ferguson, R., Macfadyen, L. P., Clow, D., Tynan, B., Alexander, S. & Dawson, S. (in press). Setting learning analytics in context: Overcoming the barriers to large–scale adoption. Invited paper for the *Journal of Learning Analytics* based on a paper presented at the 2014 Learning Analytics and Knowledge, Indianapolis, IN.

Fritz, J. (2011). Classroom walls that talk: Using online course activity data of successful students to raise self–awareness of underperforming peers. *The Internet and Higher Education, 14*(2), 89–97.

Glahn, C. (2013). Using the ADL experience API for mobile learning, sensing, informing, encouraging, orchestrating. In *International conference on next generation mobile apps, services and technologies* (pp. 268–273). New York, NY: IEEE Press.

Goldstein, P. J., & Katz, R. N. (2005). *Academic analytics: The uses of management information and technology in higher education* (Vol. 8). Louisville, CO: EDUCAUSE Center for Applied Research.

Gupta, A., & Anish, S. (2009). Insights from complexity theory: Understanding organisations better. *IIMB Management Review*. http://tejas.iimb.ac.in/articles/12.php

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

Haynes, P. (2003). *Managing complexity in the public services*. Maidenhead, UK: Open University Press.

Head, B. W., & Alford, J. (2013). Wicked problems: Implications for public policy and management. *Administration & Society*. doi:10.1177/0095399713481601

Jacob, B. (2005). Accountability, incentives and behavior: Evidence from school reform in Chicago. *Journal of Public Economics, 89*(5–6), 761–796.

Kavanagh, M. H., & Ashkanasy, N. M. (2006). The impact of leadership and change management strategy on organizationsal culture and individual acceptance of change during a merger. *British Journal of Management, 17,* S81–S103.

Kingdon, J. (1995). *Agendas, alternatives and public policies* (2nd ed.). Boston, MA: Longman.

Kotler, P., & Zaltman, G. (1971). Social marketing: An approach to planned social change. *Journal of Marketing, 35*, 3–12.

Kotter, J. P. (1996). *Leading change*. Boston, MA: Harvard Business School Press.

Kotter, J. P., & Cohen, D. S. (2002). *The heart of change*. Boston, MA: Harvard Business School Press.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How differences in recognizing one's own incompetence lead to inflated self–assessments. *Journal of Personality and Social Psychology, 77*(6), 1121–1134.

Macfadyen, L. P., & 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.

MacLennan, B. (2007). Evolutionary psychology, complex systems, and social theory. *Soundings: An Interdisciplinary Journal, 90*(3/4), 169–189.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition and productivity. McKinsey Global Institute. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

McDonnell, L. M. (1994). Assessment policy as persuasion and regulation. *American Journal of Education, 102*(4), 394–320.

McIntosh, N. E. (1979). Barriers to implementing research in higher education. *Studies in Higher Education, 4*(1), 77–86.

Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. *Research & Practice in Assessment, 9*(2), 70-89.

Mitleton–Kelly, E. (2003). Ten principles of complexity & enabling infrastructures. In E. Mitleton–Kelly (Ed.), *Complex systems & evolutionary perspectives of organisations: The application of complexity theory to organisations* (pp. 23–50). Philadelphia, PA: Elsevier.

Nakamura, R. T. (1987). The textbook policy process and implementation research. *Review of Policy Research, 7*(1), 142–154.

Norris, D., & Baer, L. L. (2013). *Building organizational capacity for analytics*. Louisville, CO: EDUCAUSE. https://net.educause.edu/ir/library/pdf/PUB9012.pdf

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), 42–67.

Pintrich, P. R. (1999). The role of motivation in promoting and sustaining self–regulated learning. *International Journal of Educational Research, 31*(6), 459–470.

Rein, M. (1976). *Social science and public policy*. Harmondsworth, UK: Penguin.

Richardson, J. (2004). Methodological issues in questionnaire–based research on student learning in higher education. *Educational Psychology Review, 16*(4), 347–358.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155–169.

Sabatier, P. A. (Ed.). (2007). *Theories of the policy process* (2nd ed.). Boulder, CO: Westview Press.

Schon, D. A., & Rein, M. (1994). *Frame reflection: Toward the resolution of intractable policy controversies*. New York, NY: Basic Books.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. doi:10.1177/0002764213498851

Siemens, G., Dawson, S., & Lynch, G. (2013). *Improving the productivity of the higher education sector: Policy and strategy for systems–level deployment of learning analytics*. Sydney, Australia: Society for Learning Analytics Research for the Australian Government Office for Learning and Teaching. http://solaresearch.org/Policy_Strategy_Analytics.pdf

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529.

Tanes, Z., Arnold, K. E., King, A. S., & Remnet, M. A. (2011). Using signals for appropriate feedback: Perceptions and practices. *Computers & Education, 57*(4), 2414–2422. http://www.itap.purdue.edu/learning/docs/research/1–s2.0–S0360131511001229–main.pdf

Thomas, H. (2010). Learning spaces, learning environments and the dis'placement' of learning. *British Journal of Educational Technology, 41*(3), 502–511.

Tiesman, G. R., van Buuren, A., & Gerrits, L. (Eds.). (2009). *Managing complex governance systems*. London, UK: Routledge.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*. doi: 10.1177/0002764213479363

Wall, A. F., Hursh, D., & Rodgers, J. W. III. (2014). Assessment for whom: Repositioning higher education assessment as an ethical and value–focused social practice. *Research and Practice in Assessment, 9*(1), 5–17.

Wiliam, D. (2010). The role of formative assessment in effective learning environments. In H. Dumont, D. Istance, & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice* (pp. 135–155): Paris, France: OECD Publishing.

Winne, P. H. (2010). Improving measurements of self–regulated learning. *Educational Psychologist, 45*(4), 267–276.

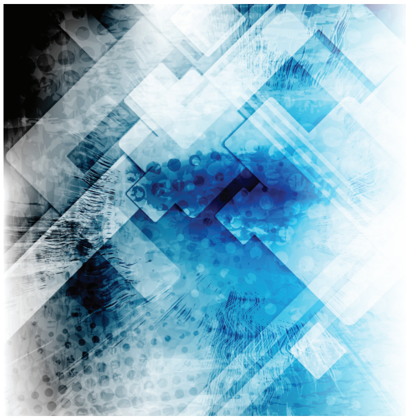
Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In A. Pardo & S. D. Teasley (Eds.), *Proceedings of the International Conference on Learning Analytics and Knowledge*. New York, NY: ACM Press.

Yanosky, R. (2009). *Institutional data management in higher education*. Boulder, CO: EDUCAUSE Center for Applied Research.

Young, J., & Mendizabal, E. (2009). Helping researchers become policy entrepreneurs: How to develop engagement strategies for evidence–based policy–making. (Briefing Paper) ODI briefing Papers. London, UK.

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self–reported versus traced goal orientation. *Learning and Instruction, 22*(6), 413–419.

Zimmerman, B. J. (2002). Becoming a self–regulated learner: An overview. *Theory Into Practice, 41*(2), 64–70. doi: 10.1207/s15430421tip4102_2



Abstract

Because MOOCs bring big data to the forefront, they confront learning science with technology challenges. We describe an agenda for developing technology that enables MOOC analytics. Such an agenda needs to efficiently address the detailed, low level, high volume nature of MOOC data. It also needs to help exploit the data's capacity to reveal, in detail, how students behave and how learning takes place. We chart an agenda that starts with data standardization. It identifies crowd sourcing as a means to speed up data analysis of forum data or predictive analytics of student behavior. It also points to open source platforms that allow software to be shared and visualization analytics to be discussed.

AUTHORS

Una–May O'Reilly, Ph.D.
Massachusetts Institute of Technology

Kalyan Veeramachaneni, Ph.D.
Massachusetts Institute of Technology

Technology for Mining the Big Data of MOOCs

Massive Open Online Courses (MOOCs) are college courses offered on the Internet. Lectures are conveyed by videos, textbooks are digitized, and problem sets, quizzes and practice questions are web–based. Students communicate with one another and faculty via discussion forums. Grading, albeit constrained by somewhat restrictive assessment design, is automated.

The popularity of MOOCs has made a high volume of learner data available for analytic purposes. Some MOOC data is just like that which comes from the classroom. This can include teaching material, student demographics and background data, enrollment information, assessment scores and grades. But very important differences arise between MOOC and classroom in how behavioral data is collected and what is observable. The platform records, unobtrusively, through input, capture every mouse click, video player control use, and every submission to the platform such as problem solution choice selection, solution composition or text entry for a forum discussion. The level of recorded detail of behavior in a MOOC vastly surpasses that recorded in conventional settings.

Very directly, this data can provide a count of problem attempts and video replays. It can reveal how long a student stayed on a textbook page or the presence of very short, quick patterns of resource consultation. It can inform an individualized or aggregated portrait of how a student solves problems or accesses resources. It presents opportunities to identify and compare different cohorts of students in significant quantities, thus enabling us to personalize how content is delivered. It allows us to study learner activities not exclusive to problem-solving, such as forum interactions and video-watching habits (Thille et al., 2014). It also facilitates predictive analytics based on modeling and machine learning.

CORRESPONDENCE

Email
unamay@csail.mit.edu

This data also contains large samples. Large sample sizes enable us to rigorously confirm or deny long held hypotheses about how learning takes place, whether there exist learning styles, whether there are effective ways to learn or teach types of material or whether there are effective concept correction strategies to help a student who has made an error.

Beyond comparative studies, from a predictive modeling standpoint, we can build and validate predictive models at a scale never done before. For example, we can now build a reliable predictor for which students will exit the course before completion (Taylor, Veeramachaneni, & O'Reilly, 2014). In short, MOOC big data is a gold mine for analytics.

The enormous potential of MOOC big data prompts the questions: what are the appropriate ways to fully tap into it? What technology can be brought to practice to analyze it more efficiently and broadly? The process of answering these questions reveals challenges. The data is *high volume* and *low–level* in nature. Complete answers to any research question need to analyze the data from multiple entities, i.e., courses, platforms, institutions. The perspectives of multiple parties – students, instructors and education researchers – need to be explored.

We have decided to focus our research agenda on the challenges that arise from MOOC data characteristics and analytics needs. We have embraced increasing the number of contributors to MOOC analytics and accelerating analytics accomplishments as our central mission. We are focusing on developing community–oriented means of sharing software and analytic development efforts.

We start by proposing data standardization as a cornerstone. It will resolve the different formats of data resulting from different platforms. It will prevent MOOC data from following the path of healthcare data, which, even if privacy issues are completely resolved, is fragmented by different formats. It will also make the task of extracting variables for analyses more efficient, collaborative and sharable. We next propose easy–to–use, web–based platforms that democratize different aspects of data analytics:

- *MOOCviz* lets anyone share visualization software and their analytic renderings.
- *FeatureFactory* helps learning scientists enumerate possible variables for their models.
- *LabelMe–Text* helps learning scientists engage the crowd to get help tagging forum posts before they use machine learning to automate a labeler from the tagged examples.

MOOCdb – A Cornerstone for Shared Analytics

Large sample sizes enable us to rigorously confirm or deny long held hypotheses about how learning takes place, whether there exist learning styles, whether there are effective ways to learn or teach types of material or whether there are effective concept correction strategies to help a student who has made an error.

In order for a data oriented platform or framework to allow anyone to use it, it needs to either deal with many formats of data or be able to expect that all data is in a common format. The former proposition imposes a lot of extra work versus the latter. It leads to different versions of software. It bulks logic in software to dealing with format differences and it requires software updates every time a new format emerges. Thus, to make the latter proposition viable, we have pioneered a standardized schema for MOOC data (i.e., a data model) that is platform agnostic. It is called MOOCdb (Veeramachaneni, Halawa, et al., 2014).

The MOOCdb data model originally organized MITx data generated from the MITx platform that has now transitioned to edX. It offers advantages beyond what we emphasize here, among them removing the need to share data, independence from platform specifics and facilitating a data description that outsiders can refer to when contributing expertise in data privacy protection or database optimization. During the past year, we have adapted it to also capture the data subtleties and idiosyncrasies of both edX and Coursera platforms. A periodically updated technical report explains the data model, all the fields and how they are assembled for each platform. Complete documentation for MOOCdb and its data model will be perpetually updated via the wiki site http://moocdb.csail.mit.edu/wiki.

The MOOCdb data model is based on some basic core actions that students take on any online learning platform. Students usually interact with the platform in four different modes: *Observing, submitting, collaborating* and *giving feedback*. In observing mode students are simply browsing the online platform, watching videos, reading material, reading books or watching forums. In submitting mode, students submit information to the platform. This includes submissions towards quizzes, homework, or any assessment modules. In collaborating mode students interact with other students or instructors on forums, collaboratively editing wiki or chatting on Google hangout or other hangout venues (Veeramachaneni, Halawa, et al., 2014).

To date, much of the analyses on MOOC data have been conducted with techniques transferred from conventional learning analytics or modestly adapted from them.[1] In the first three stages of their study, Breslow et al. (2013) followed a conventional methodology adapted for MOOC relevant questions. They worked with coarse–grained variables. That is, they studied the aggregate of certificate earners (choosing not to further subdivide students), they operationalized achievement to use the course grade (choosing not to consider specific problem set grades or time sequences of assessment grades) and they referenced factors such as age, gender and highest degree earned (choosing not to reference behavioral factors such as instructional component access). MOOCdb standardization will further leverage such work because it supports the extraction of quantities that can be composed into fine grained variables. It allows anyone to formulate (and answer) learning science research questions that are adaptations of conventional methods considering finely subdivided students, their achievements and their access of MOOC's instructional components.

Infrastructure for Sharing Data Visualizations

Transforming data into meaningful visualizations is a core part of any data science. In MOOC data science, different institutions, local research communities, user groups and other sorts of organizations, each have multiple stakeholders who have different needs that require data to be transformed in a different way and visualized. Ideally, they want to support each other as much as possible in this context by sharing software, demonstrations and opinions on design and interpretations of data.

Visualization infrastructure can provide one means of supporting this. HarvardX and MIT's Office of Digital Learning enable visualizations of their MOOC data[2,3] via complementary website entitled *Insights*. These visualizations use world maps to show enrollment, certificate attainment by country, gender, education levels and age composition (Ho et al., 2014; Nesterko et al., 2013). Visualizations referencing clickstream or forum data are currently not available[4] likely because plotting these streams is significantly more complicated. A streamlined workflow that reduces development time through software sharing and data standardization would reduce these complications.

The *Insights* website is also used as a distribution point and makes a modest attempt to encourage other visualizations that reference the data. For example, along with the data that populate visualizations, *Insights* makes source code and documentation available for download,[5] though only as separate, non–integrated files. The website exemplifies a strong but minimal starting point for providing visualization infrastructure. Ideally, even beyond supporting better–integrated software sharing, an infrastructure needs to support the contribution of new visualizations. These should be able to come from others, i.e., not only the site's creators. Opening access to the community, so they can contribute, will allow many different questions to be answered by data visualizations expressed in multiple ways. It will address the reality that different people perceive different visualizations as useful.

People analyzing visualizations for their usefulness tend to zero in on either on the aesthetics of the visualization, e.g., a style choice like bar or pie chart, color or interaction

We start by proposing data standardization as a cornerstone. It will resolve the different formats of data resulting from different platforms. It will prevent MOOC data from following the path of healthcare data, which, even if privacy issues are completely resolved, is fragmented by different formats.

[1] In the first paper in RPA on MOOCs, Breslow et al. (2013) note: Our first challenge has been choosing, or in some cases adapting, the methodological approaches that can be used to analyze the data. If educational researchers studying conventional brick and mortar classrooms struggle to operationalize variables like attrition and achievement, it is doubly difficult to do so for MOOCs (p. 14).

[2] MITx Insights is a collection of interactive data visualizations for all MITx offerings, updating at frequent, regular intervals. These visualizations are released along side a complementary set of visualizations from the HarvardX Research Committee. (url: http://odl.mit.edu/insights/)

[3] HarvardX Insights is a collection of interactive visualizations of learner data, which dynamically update at frequent, regular intervals. (url: http://harvardx.harvard.edu/harvardx–insights)

[4] In their reporting, the team notes: "The MITx and HarvardX Research teams intend for future interactive visualizations to include more nuanced descriptions of student participation and learning in our open online learning environments."

[5] It is highly structured and organized so whether it will support different visualizations is an open question (see e–literate for an opinion).

mode, or on the way the data was organized and aggregated before it was visualized. Such remarks motivate a fundamental goal for visualization infrastructure: to support a proliferation of many views of same data. This goal has driven us to develop a platform called MOOCviz that we now describe.

For example, we can now build a reliable predictor for which students will exit the course before completion.

MOOCviz – Sharing Software and Outcomes of Visualization

The *MOOCviz* platform (Figure 1) is designed to serve the diverse needs of a broad group of stakeholders and facilitates the sharing of software, demonstrations and opinions on design and interpretations of data. It enforces source code organization, allows source code to be contributed to a repository and it provides a means of web–based discussion around a visualization, all fundamental tenets for a community oriented infrastructure.

Transforming data to create visualization typically requires three steps: *source data extraction, variable formation (typically aggregation)* and *rendering*. Each of these steps is somewhat specialized according to each situation. They embed some assumptions and integrate some heuristics to transform and shape the data to create an interesting and informative visualization. Anyone with access to MOOC data in MOOCdb schema can develop a brand new visualization, modularize their software into the aforementioned three steps, extract, aggregate and render, and then upload the modules into *MOOCviz*'s software repository along with their first demonstration of the new visualization for other members to use and view.

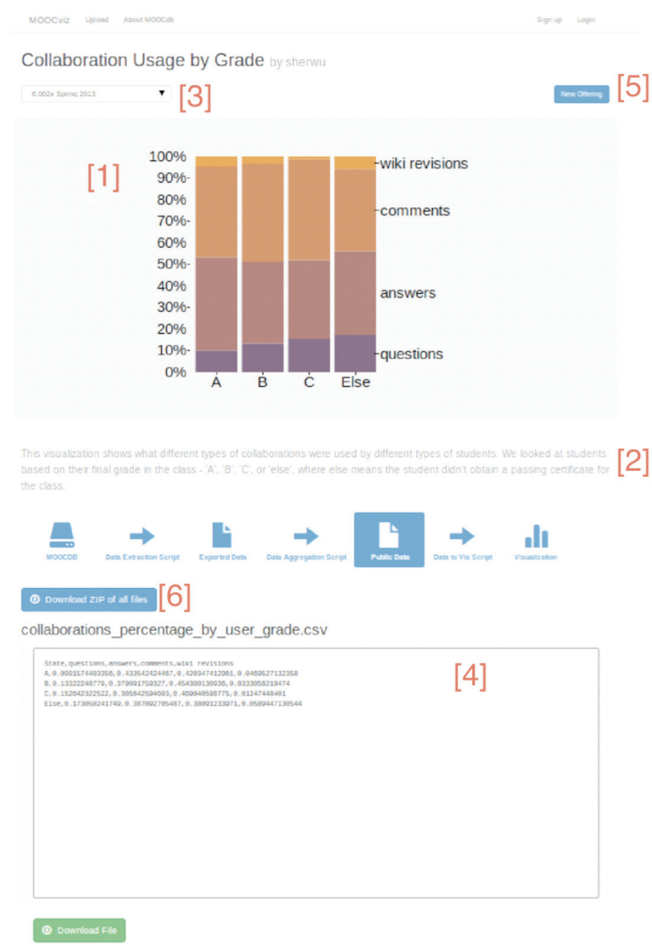


Figure 1. Current state of the *MOOCviz* platform. Users can select the course for which they would like to see the visualization (see [3]). The visualization is rendered in panel [1] and is described below the panel (see [2]). The workflow that generated the visualization from MOOCdb is shown below the description. users can click on any of the icons in the workflow and corresponding software or data is shown in panel parked as [4]. Users can upload the visualization for a new course by using the "New Offering" functionality (see [5]). [6] allows usersto download the entire code from data extraction, aggregation to visualiztion.

[6] In order to help a viewer choose between different visualizations, it will use popularity to rank multiple visualizations and only show the most popular one.

The *MOOCviz* platform software will eventually be shared under an open source license, and an organization or an instructor will be able to download and install it to create an independent instance, which they can populate with visualizations of their own data in MOOCdb format. Any member of the community will be able to enhance the platform's open source software and customize it to support specific use cases; e.g., cross–course comparisons or a single course report with multiple visualizations.

A *MOOCviz* platform offers:

- A central, shared gallery of participant–generated visualizations for a list of courses for which they have been rendered.
- The ability for the participants to download the software that generates visualizations and execute it over their own course data that is formatted in MOOCdb schema. They will also be able to automatically package the resulting rendered visualization and upload it to the gallery, adding to the list of courses.
- A means to contribute software for new visualizations to the gallery via the *MOOCviz* web–based interface.
- A means of commenting on any existing visualization by posting in the *comments* section underneath it. Discussions are free form. They likely will extend beyond the interpretation or thoughts provoked by the visualization to the ways that the data have been transformed in extraction and aggregate steps. We expect that discussions will stimulate ideas for new visualizations.

Infrastructure for Supporting Feature Engineering

In order for a data oriented platform or framework to allow anyone to use it, it needs to either deal with many formats of data or be able to expect that all data is in a common format.

Scaling feature engineering involves three processes: proliferation of an *ideation* process, the process in which candidate features are posited; support for an *operationalization* process, in which a mapping is formed between the data sources and the feature; and a feature *extraction* process, in which software is written to realize instances of these features.

The study of stopout, that is, predicting when students stop engaging course material before completion, provides an example (Taylor et al., 2014). If the outcome set is whether or not a student stops out, what predicts a stopout could include frequency of forum posts, grades to date, most recent problem set score, time spent watching videos, etc.

We have been formulating predictive and explanatory features for stopout. In the course of doing so, we have observed that the set of possible features for an outcome is likely much larger than we ourselves can propose (Veeramachaneni, O'Reilly, & Taylor, 2014). This is because our own experiences (or lack thereof), biases and intellectual context can go only so far and may be imposing limits on our investigations. This is a shortcoming not unique to us alone.

When working on stopout prediction (Taylor et al., 2014), we first tried to address this shortcoming by setting up meetings with students and instructors of a MOOC. At the meeting, we would solicit in person via a somewhat informal protocol, a group's input for predictors of stopout. We asked our helpers to fill out a form listing variables that would predict a student stopping out. We would then operationalize these variables via extraction and some modest arithmetic and add them to our predictor set (Veeramachaneni, O'Reilly, et al., 2014).). These exercises begged a general question: how can any MOOC data science group access a wider swath of the MOOC community to expand their feature/predictor list? As well, considering our mission to enable technology for MOOC analytics, how can we provide a general means of crowd access to the MOOC data science community at large?

FeatureFactory – Engaging the MOOC Crowd to Provide Hypotheses

To address both these questions, we are developing a second web–based collaborative platform called *FeatureFactory*. Our current version of this platform is shown in Figure 2. *FeatureFactory* offers two modes of engagement:

- The *solicit* mode is used by MOOC data science, education technology, or learning science research teams. A team describes the outcome it is currently studying or trying to predict. They give examples of what features or explanations are sought and it solicits help from the MOOC crowd.
- In the second mode, *helping*, the crowd proposes, explanations or variables, and suggests means to operationalize them. They provide comments on proposal or vote them up or down in popularity. The software savvy among them write and share software scripts written to operationalize the most popular or compelling proposals.

The MOOCdb data model is based on some basic core actions that students take on any online learning platform. Students usually interact with the platform in four different modes: Observing, submitting, collaborating and giving feedback.

Like *MOOCviz*, we intend to open source license and share the *FeatureFactory* platform software, so that an organization can create its own independent instance for local use. An organization can also customize their instance by modifying the platform source. They can use their platform in contexts when they need to garner assistance from the MOOC crowd.

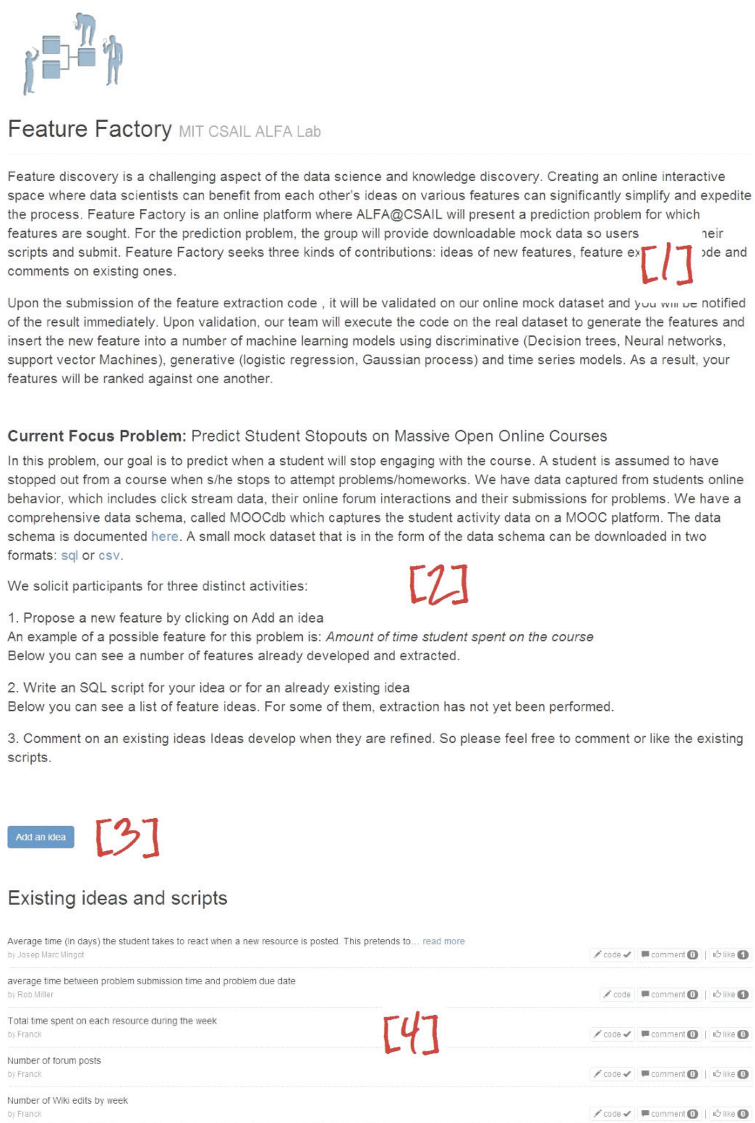


Figure 2. Current state of the *FeatureFactory* platform. In this illustration we show a screen shot of the website. First the rationale behind the *FeatureFactory* is described (see [1]), the current prediction problem of interest is described and the role participants can play is described (see [2]). Participants can submit a new idea using "Add an idea" (see [3]). Ideas collected so far are revealed under "Existing ideas and scripts" (see [4]). Participants can view the code (if available), comment on the idea and vote on the idea. All input from participants is collected in the back end in a database.

Infrastructure for Annotating Text

A central component of MOOC data is discussion forums. They are of great interest because they provide a lens on inter–student communication that, in turn, relates to learning science theories of engagement and achievement and self–efficacy. Most such language understanding tools rely on annotations of the content by humans (Gillani, 2013; Gillani & Eynon, 2014) and then employing machine learning to automatically annotate the text. The annotations range from qualifying the sentiment of the post, to tagging the posts by their types (content related, social affective, administrative, and other) to type of post (help seeking, help providing, neither) and many others. These tags help analyze the posts to understand the mood of the class, group posts by categories when presenting to the instructors, teaching assistants and others, categorizing students based on their post types so interventions can be designed, generating predictive variables for models on a per student basis and understanding the social discourse in the course (Rosé et al., 2014; Yang, Sinha, Adamson, & Rosé, 2013).

A working paper by Stump, DeBoer, Whittinghill, and Breslow (2013) provides a detailed account of how a protocol to annotate MOOC discussion forum posts was developed. The authors employed two students and used themselves to annotate the posts using a pre–determined set of labels derived from a categorization scheme. To facilitate their workflow they passed around an encrypted csv file that recorded labels. They then evaluated the quality of human annotations via a number of metrics that relate to inter–rater reliability. They finally filtered out ambiguously labeled posts. While they had over 90,000 forum posts, they found it impossible to examine and label all of them. They had to settle for referencing ~4,500 labeled posts. It is obvious that interpreting an entire set of posts would be preferable. But the process is slowed by the involvement of humans and hindered by the awkwardness of an ad hoc workflow. Concurrently, discussion arose outside the project arguing for an alternative annotation scheme (Gillani, 2013; Gillani & Eynon, 2014). This implies that annotation needs to become much easier because it will need to be done many ways by multiple research teams.

MOOCdb standardization will further leverage such work because it supports the extraction of quantities that can be composed into fine grained variables.

This context led us to consider what MOOC specific technology we could design to deal with such a large scale set of text and to support labeling according to the different annotation schemes of different studies. First, a web–based framework can support crowd based labeling for larger scale labeling. Second, the process and the workflow for processing labels can be streamlined. Third, much of the labeling can be automated. Machine learning can be used on the set of labeled posts to learn a rule for labeling the others, based upon features in the post. To address these needs, we are developing a web–based platform called *Label Me–Text*.

LabelMe–Text – Engaging the MOOC Crowd to Help with Forum Annotation

We developed an online platform where users would post their tagging projects and a crowd of helpers can participate in MOOC data science by selecting a project and tagging the content based on some instructions. We call the online collaborative platform that serves this purpose *LabelMe–Text*'s.[7] *LabelMe*'s current incarnation is shown in Figure 3. It works in the following ways:

Transforming data into meaningful visualizations is a core part of any data science.

- Users requiring annotation of natural language can create an annotation project by providing a csv file for the content, instructions and examples for tagging.
- Taggers (*LabelMe–Text*'s crowd) can participate by selecting a project, following the instructions and tagging the content.
- A database consisting of tags for the content for the project is initialized and populated as taggers work. A number of analytic services are provided around this database such as evaluation of inter rater reliability, summary of tags, and summary of activity for a project (how many taggers helped, time series of number of tags).

[7] A framework called LabelMe already exists in the Computer Vision community (Russell, Torralba, Murphy, & Freeman, 2007). We used the same name, but identify it with suffix – text, by calling it LabelMe–Text.

- A service can be called upon to filter the tagged data based on the reliability measures just mentioned. It then uses methods based upon latent semantic analysis to learn a tagging model.
- Taggers (*LabelMe–Text*'s crowd) are given credit for every tag they have provided and the number of their tags that pass the filters to be used in model learning.

Transforming data to create visualization typically requires three steps: source data extraction, variable formation (typically aggregation) and rendering.

Like *MOOCviz* and *FeatureFactory*, *LabelMe–Text* is open source software. Its eventual release will to support organizations that wish to download and create a local version of if for internal use.

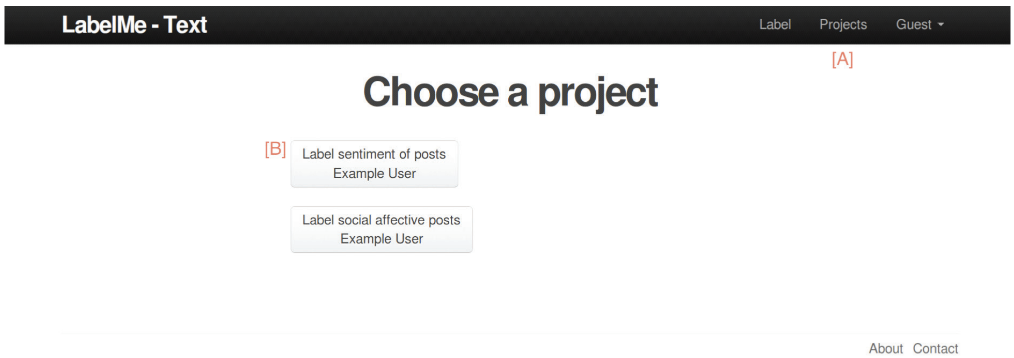


Figure 3. Crowd can select a project posted by a researcher by clickin on "Projects" marked using [B]. In this screen shot two such projects appear where it is marked as [A].

A central component of MOOC data is discussion forums. They are of great interest because they provide a lens on inter–student communication that, in turn, relates to learning science theories of engagement and achievement and self–efficacy.

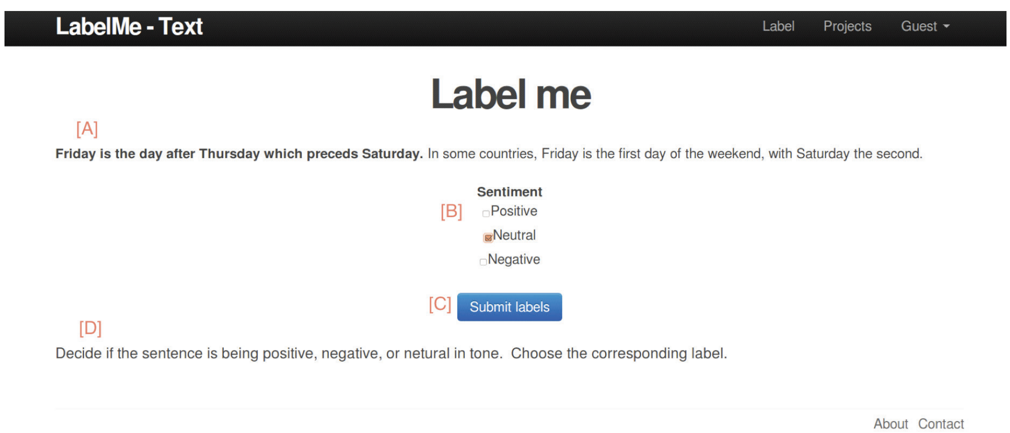


Figure 4. Once users select the project, they then proced to tagging/annotating a post/sentence dynamically selected by the platform from the pool of posts/sentences that need to be tagged. The sentence is displayed (see [A]), the choices for tags are displayed underneath it (see [B]) and instructions for tagging are presented as well (see [D]). The user can select the tag and hit "Submit Labels" (see [C]). All inputs from the participants/users are stored in a structured format in the back end in a database.

Conclusion

This paper considers the complexity MOOCs bring into learning science in view of the novel nature of the data they collect. It identifies certain technology challenges that need to be resolved before we can exploit the big data in MOOCs to its full potential. We call for enabling technology and for setting a course towards standardization and web–based platforms that help a large community of people to collaborate on developing analytics. We advocate frameworks that are deliberately open source so that, when they are released, everyone will be able to customize, refine and advance them.

AUTHORS NOTE:
We would like to thank the following student teams and students who, by contributing substantial efforts toward platform development, have been helping us fulfill our vision as described in this paper: Kiarash Adl (*FeatureFactory*), Preston Thompson, Colin Taylor, Brian Bell, Sherwin Wu (*MOOCviz*), Roy Wedge (*LabelMe*), Franck Dernoncourt (*MOOCdb*), Quentin Agren (*MOOCdb*), Sherif Halawa (*MOOCdb*). We would also like to acknowledge discussions with Juho Kim. We are grateful for funding from Quanta Computer. We also thank our reviewers.

References

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*(1), 13–25.

Gillani, N. (2013). Learner communications in massively open online courses. *OxCHEPS Occasional Paper, 53*, 1–51.

Gillani, N., & Eynon, R. (2014). Communication patterns in massively open online courses. *The Internet and Higher Education, 23*, 18–26.

Ho, A. D., Reich, J., Nesterko, S., Seaton, D. T., Mullaney, T., Waldo, J., & Chuang, I. (2014). HarvardX and MITx: The first year of open online courses (HarvardX Working Paper No. 1).

Nesterko, S. O., Dotsenko, S., Han, Q., Seaton, D., Reich, J., Chuang, I., & Ho, A. D. (2013). Evaluating the geographic data in MOOCs. In *Neural Information Processing Systems*.

Rosé, C. P., Carlson, R., Yang, D., Wen, M., Resnick, L., Goldman, P., & Sherer, J. (2014, March). Social factors that contribute to attrition in MOOCs. In *Proceedings of the first ACM conference on Learning@ scale conference* (pp. 197–198). ACM.

Russell, B., Torralba, A., Murphy, K., & Freeman, W. T. (2007). LabelMe: A database and web–based tool for image annotation. *International Journal of Computer Vision, 77*(1–3), 157–173.

Stump, G. S., DeBoer, J., Whittinghill, J., & Breslow, L. (2013). Development of a framework to classify MOOC discussion forum posts: Methodology and challenges. TLL Working Paper, 2013.

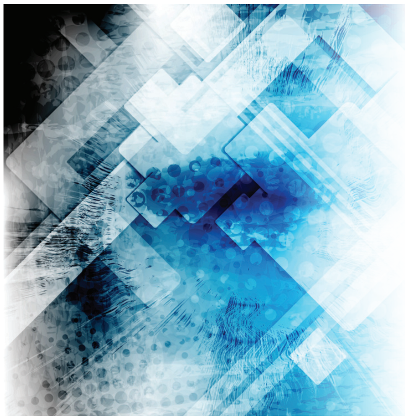
Taylor, C., Veeramachaneni, K., & O'Reilly, U. M. (2014). Likely to stop? Predicting stopout in massive open online courses. *arXiv preprint arXiv:1408.3382.*

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*(2), 5-16.

Veeramachaneni, K., Halawa, S., Dernoncourt, F., O'Reilly, U. M., Taylor, C., & Do, C. (2014). MOOCdb: Developing standards and systems to support MOOC data science. *arXiv preprint arXiv:1406.2015.*

Veeramachaneni, K., O'Reilly, U. M., & Taylor, C. (2014). Towards feature engineering at scale for data from massive open online courses. *arXiv preprint arXiv:1407.5238.*

Yang, D., Sinha, T., Adamson, D., & Rosé, C. P. (2013). Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data–Driven Education Workshop*.



Abstract

Many university leaders and faculty have the goal of promoting learning that connects across domains and prepares students with skills for their whole lives. However, as assessment emerges in higher education, many assessments focus on knowledge and skills that are specific to a single domain. Reworking assessment in higher education to focus on more robust learning is an important step towards making assessment match the goals of the context where it is being applied. In particular, assessment should focus on whether learning is robust (Koedinger, Corbett, & Perfetti, 2012), whether learning occurs in a way that transfers, prepares students for future learning, and is retained over time; and also on skills and meta–competencies that generalize across domains. By doing so, we can measure the outcomes that we as educators want to create, and increase the chance that our assessments help us to improve the outcomes we wish to create. In this article, we discuss and compare both traditional test–based methods for assessing robust learning, and new ways of inferring robustness of learning while the learning itself is occurring, comparing the methods within the domain of college genetics.

AUTHORS

Ryan S. Baker, Ph.D.
Teachers College, Columbia University

Albert T. Corbett, Ph.D.
Carnegie Mellon University

Assessment of Robust Learning with Educational Data Mining

In recent years, the historical monopoly of universities in higher education has been challenged by new entrants, including for–profit universities and massive online open courses (Hanna, 1998; Vardi, 2012). This change has brought to the forefront questions about what the core goals of higher education are: Is it to train a workforce in specific employable skills (Sperling & Tucker, 1997)? Or is it to promote learning that connects across domains and prepares students to learn the new skills and disciplines that emerge during their years in the workforce (Knapper & Croppley, 2000)? To put it another way, is the goal of higher education to learn competencies, or to learn meta–competencies which cut across domains (e.g., Buckingham Shum & Deakin Crick, 2012)?

While much of the learning that goes on in higher education pertains primarily to the content area of the class being taken, students can learn in a specific fashion or in a more general fashion. Increasingly, researchers in the learning sciences have presented evidence that it is possible to measure whether learning is *robust* – defined in Koedinger et al. (2012) as learning that can transfer to related situations (Fong & Nisbett, 1991; Singley & Anderson, 1989), prepares students for future learning (Bransford & Schwartz, 1999; Schwartz & Martin, 2004), and is retained over the long–term (Bahrick, Bahrick, Bahrick & Bahrick, 1993; Schmidt & Bjork, 1992).

CORRESPONDENCE

Email
ryanshaunbaker@gmail.com

To the extent that creating more robust learning is the primary goal of higher education, the way assessment is used may need to change. While some argue for a switch to self–assessment (e.g., Boud & Falchikov, 2006), we still see a need for instructor and curriculum–led assessment. But there is a challenge for those developing assessments for higher education; it is much easier to measure didactic knowledge or concrete skill than to measure the type of learning that has been argued for.

Nonetheless, whether learning is robust can be measured. Paper tests measuring retention and transfer have been in use for quite some time (cf. Gick & Holyoak, 1983; Surber & Anderson, 1975), with paper tests measuring a student's preparation for future learning (PFL) emerging about a decade ago (Bransford & Schwartz, 1999; Schwartz &

Martin, 2004). In this article, we discuss examples of this work within the domain of college genetics. Increasingly, it is also a goal of assessment in higher education to measure skills that cut across domains, such as science inquiry and help seeking (cf. Puncochar & Klett, 2013), and to measure robust learning of these skills while learning is ongoing (cf. Linn & Chiu, 2011). To this end, we will also discuss measures of robust learning that can measure robust learning of domain content, but also domain–general skills, in a fashion that is integrated into instruction. We discuss these new forms of assessment in terms of the same domain of college genetics for understandability; but, as we will discuss, many of the new forms of assessment are potentially meaningful domain–general.

These new forms of assessment are based on the emerging methods of educational data mining (EDM; Baker & Siemens, in press; Baker & Yacef, 2009; Romero & Ventura, 2007). Within educational data mining, the voluminous data increasingly becoming available to learners, particularly from online learning environments, becomes a source of information that can be used to identify complex learning behaviors and ill–defined or complex skill (cf. Kinnebrew & Biswas, 2012; Sao Pedro, Baker, & Gobert, 2012). These data are sometimes analyzed by use of knowledge engineering methods, where research analysts identify meaningful patterns in data by hand (e.g., Aleven, McLaren, Roll, & Koedinger, 2006), and is sometimes analyzed using automated methods such as sequence mining (Kinnebrew & Biswas, 2012) or classification (Sao Pedro et al., 2012). While knowledge engineering can be similar to traditional psychometric approaches for assessment development such as evidence–centered design (Mislevy, Almond, & Lukas, 2004), and advanced ECD–based models of complex student skill can resemble EDM models developed using automated discovery (see Shute & Ventura, 2013 for examples), the development methods of EDM and ECD differ, as do their validation. Educational data mining methods are often validated by developing the models on one set of students and testing them on another; some EDM methods are also validated on data from new domains or contexts (Sao Pedro, Gobert, & Baker, 2014) or data from new learner populations (Ocumpaugh, Baker, Kamarainen, & Metcalf, 2014). In addition, EDM–based assessments are typically validated for agreement with human judgments about a construct's presence which themselves are known to be reliable (Ocumpaugh et al., 2014; Sao Pedro et al., 2014), and are based on data features thought by domain experts to be plausibly related to the construct of interest (Sao Pedro et al., 2012). In some cases, their internal structure is not considered in detail, being too complex for a human analyst to understand without hours of study, but that is not true of all EDM–developed models; the models resulting from the EDM process are particularly simple for the cases presented in this paper. A full discussion of educational data mining methods is outside the scope of this paper, but richer summaries are provided in the papers (Baker & Siemens, in press; Baker & Yacef, 2009; O'Reilly & Veeramachaneni, 2014; Romero & Ventura, 2007) and the textbook (Baker, 2013).

To put it another way, is the goal of higher education to learn competencies, or to learn meta–competencies which cut across domains?

EDM–based assessment has multiple benefits compared to traditional methods of assessment: If the models are designed appropriately, they can be used in real time to make assessment during learning and support real time intervention. In addition, since the models typically make inferences based on ongoing interaction between a student and online system, they can replicate the assessments made by more traditional instruments without needing to take the student's time up with a paper test. See, for instance, Feng, Heffernan, and Koedinger (2009), who show that EDM models based on student interaction can accurately predict standardized exam scores.

Case Study in College Genetics Tutor

In this article, we discuss the potential for assessment of robust learning in higher education, both with traditional methods and educational data mining methods, using examples drawn from the domain of genetics. Genetics is an important topic because it is a central, unifying theme of modern biology and because it provides the foundation for many advances in 21st century technology. It is a challenging topic for students, because it depends heavily on problem solving (Smith, 1988). Finally, it is a relevant topic because it affords an interesting form of superficial learning: Students can develop successful problem solving algorithms that are not well grounded in the underlying biology.

We discuss this specifically within the context of work to develop and utilize an e–learning system for college genetics, the Genetics Cognitive Tutor (GCT; Corbett, Kauffman, MacLaren, Wagner, & Jones, 2010). GCT is focused on helping students learn not only genetics domain materials, but also the complex abductive reasoning skills needed to make inferences within this domain. Abductive reasoning skills involve reasoning "backward" from empirical observations (e.g., a daughter of unaffected parents is affected by a genetic trait) to an explanation for the observations (each parent must carry a recessive allele for the trait). Abductive reasoning skills are an important part of the undergraduate learning experience, not just in genetics, but across domains, because they are essential skills in formulating scientific knowledge, and in applying such knowledge to diagnostic tasks.

To the extent that creating more robust learning is the primary goal of higher education, the way assessment is used may need to change.

Cognitive Tutors are a type of online learning system where students complete problems (in genetics or other domains) within the context of activities designed to scaffold problem solving skill (Koedinger & Corbett, 2006). The student completes problems within an interface that makes visible cognitive steps of the problem solving process visible, and receives instant feedback on their performance. Student performance is analyzed in real time according to a cognitive model of the domain. If a student's answer indicates a known misconception, the student receives instant feedback on why their answer was incorrect. At any time, the student can request help that is sensitive to their current learning context.

GCT has more than 175 genetics problems, divided into 19 modules, which address topics in Mendelian inheritance, pedigree analysis, genetic mapping, gene regulation, and population genetics. An average of about 25 steps is needed for each of the 175 problems in GCT. It has served as supplementary instruction in a variety of undergraduate biology classes in a wide range of public and private universities in the United States and Canada (Corbett et al., 2010). It has also been used by students enrolled in high school biology classes (e.g., Corbett et al., 2013a, 2013b; Baker, Corbett, & Gowda, in press).

The goal of GCT is not just to promote immediate learning of the exact content studied within the system, but to promote robust learning as defined above. As such, research during the development of GCT focused on assessing robust learning, both after use of the system and during use of the system.

Assessing Robust Learning in College Genetics with Tests

Tests historically have been one of the most common methods for assessing robust learning. They are clearly the most straightforward way of doing so; for instance, a test can be administered immediately at the end of an activity or multiple times during the semester.

Within educational data mining, the voluminous data increasingly becoming available to learners, particularly from online learning environments, becomes a source of information that can be used to identify complex learning behaviors and ill–defined or complex skill.

The history of research on retention of material, both in research settings and classroom settings, has depended heavily on retesting the same material or same skill. This has been conducted through classical paper tests (Surber & Anderson, 1975), and in online systems such as the Automatic Reassessment and Relearning System, which retests a student on material they have learned at increasing time intervals (Wang & Heffernan, 2011).

So too, a great deal of the research on whether knowledge is transferrable has depended on paper tests, although performance–based measures have also been used in some cases (e.g., Singley & Anderson, 1989). And again, while much of the research on preparation for future learning has utilized complex learning activities and resources, the assessments have often involved paper post–tests, albeit post–tests with learning resources embedded (e.g., Bransford & Schwartz, 1999; Chin et al., 2010, Schwartz & Martin, 2004).

In several GCT studies, paper assessments of retention, transfer, and PFL were administered to study the robustness of student learning. For a selected set of lessons, transfer tests and PFL tests were administered to students immediately after they completed use of the system. For example, after students completed a lesson on 3–factor cross reasoning, they were assigned "gap filling transfer tests" (VanLehn, Jones, & Chi, 1992) where they had to complete problems for which a core case in the original formulas they learned did not apply. The problem is solvable and most of the students' problem solving knowledge directly applies; however, the student can only complete the task if they can draw on their conceptual understanding of that problem solving knowledge to fill in the gap that results from the missing group.

In the preparation for future learning tests, material beyond the current lesson was involved. For example, for the PFL test for a lesson on 3–factor cross, students were asked to solve parts of a more complex 4–factor cross problem. The reasoning is related to solving a 3–

factor cross problem, but substantially more complicated, making it unlikely that the student could discover an effective solution method during the test. Instead, the test gave the student a textual description of the solution method, and then asked them to solve the problem. For retention, the same types of problems as seen in GCT were given to students in a paper form, but one week later.

Students were generally successful on each of these tests. Student performance on the test of retention was high (M = 0.78, SD = 0.21), comparable to the immediate post–test that covered the same skills as the lesson (M = 0.81, SD = 0.18), and substantially higher than the pre–test (M = 0.31, SD = 0.18). Student performance on the PFL test (M = 0.89, SD = 0.15) and transfer test (M = 0.85, SD = 0.18) was also high, approximately equal to the immediate basic problem–solving post–test (Baker, Gowda, & Corbett, 2011a, 2011b). These results indicated that the GCT was generally successful at promoting robust learning.

It would be possible to stop at this point, and simply offer that conclusion; however, it would be useful to be able to infer the robustness of student learning earlier than after the learning episode. Beyond that, it is desirable to be able to infer the robustness of learning during the learning episode, when it is easier to intervene. In addition, tests are time consuming to administer. As such, the following sections describe our work to infer robust learning in real time, and thus these tests were used as the basis for further research.

Inferring Robust Learning in College Genetics with Learning Models

A second way to infer robust learning is through the use of automated models that infer student skill learning. This method is not specifically tailored to robust learning – it is tailored to the learning that occurs in the lesson being studied – but may be successful at predicting robust learning as well. There are examples of this type of research going back several years. For example, Jastrzembski, Gluck, and Gunzelmann (2006) have used this type of modeling to predict student retention of knowledge, within an online learning system teaching flight skills.

In this article, we discuss the potential for assessment of robust learning in higher education, both with traditional methods and educational data mining methods, using examples drawn from the domain of genetics.

Within GCT, knowledge is modeled in real time using an algorithm named Bayesian Knowledge Tracing (Corbett & Anderson, 1995). Bayesian Knowledge Tracing (BKT) is the classic algorithm for modeling student knowledge within online problem solving; it has been used in many systems and analyses, cited thousands of times, and performs comparably to or better than other algorithms for cases where its assumptions apply (see results and review in Pardos, Baker, Gowda, & Heffernan, 2011).

Bayesian Knowledge Tracing can be seen as either a simple Bayes Net or a simple Hidden Markov Model (Reye, 2004). Within BKT, a probability is continually estimated for the probability that the student knows each skill in the lesson or system. These probabilities are updated each time a student attempts a new problem solving step, with correct actions treated as evidence the student knows the skill, and incorrect actions and help requests treated as evidence that the student does not know the skill. As with psychometric models such as DINA (deterministic inputs, noisy and gate; Junker & Sijtsma, 2001), (Junker & Sijtsma, 2001), BKT takes into account the possibility that a student may have gotten a correct answer by guessing, or may have slipped and obtained an incorrect answer despite knowing the relevant skill. However, BKT does not typically account for the possibility that a student may forget what they have learned (but see an example where it is extended to do so in Qiu, Qi, Lu, Pardos, & Heffernan, 2011), or that a student may have developed shallow knowledge that will not transfer between contexts.

Bayesian Knowledge Tracing and its properties are discussed in detail in dozens of papers, with the first being Corbett and Anderson (1995). For reasons of space, only a brief description will be given here. Bayesian Knowledge Tracing calculates the probability that a student knows a specific skill at a specific time, applying four parameters within a set of equations, and repeatedly updating probability estimates based on the student's performance. This process is carried out separately for each of the cognitive skills in the domain – there are eight such skills in the case of the GCT lesson on 3–factor cross. The model makes the assumption that at each problem step, a student either knows the skill or does not know the skill. It was originally thought that the model also made the assumption that each student response will either be correct or incorrect (help requests are treated as incorrect by the model), but it has been shown more recently that extending BKT to handle probabilistic input is very easy (e.g., Sao Pedro et al., 2014). If the student does not know a specific skill, there is nonetheless a probability **G** (for "Guess") that the student answer correctly. Correspondingly, if the student does not know the skill, there is a probability **S** (for "Slip") that the student will answer incorrectly. When the student starts the lesson, each student has an initial prior probability $\mathbf{L}_0$ of knowing each skill, and each time the student encounters the skill, there is a probability **T** (for "Transition") that the student will learn the skill, whether or not they answer correctly. Each of the four parameters within Bayesian Knowledge Tracing are fit for each skill, using data on student performance; there is current debate on which method is best for fitting parameters, but several approaches seem reasonable and comparably good (see discussion in Pardos et al., 2011).

Every time the student attempts a problem step for the first time, BKT updates its estimate that the student knows the relevant skill. The procedure is as follows (the relevant equations are given in Figure 1):

> 1.) Take the probability that the student knew the skill before the current problem step $\mathbf{L}_{n-1}$ and the correctness of the student response, and re–estimate the probability that the student knew the skill before the current problem step.
>
> 2.) Estimate the probability that the student knows the skill after the current problem step, using the adjusted probability that the student knew the skill before the current problem step, and the probability **T** that the student learned the skill on the step.

$$P(L_{n-1}|Correct_n) = \frac{P(L_{n-1}) * (1 - P(S))}{P(L_{n-1}) * \left(1 - P(S)\right) + \ (1 - P(L_{n-1})) * (P(G))}$$

$$P(L_{n-1}|Incorrect_n) = \frac{P(L_{n-1}) * P(S)}{P(L_{n-1}) * P(S) + \ (1 - P(L_{n-1})) * (1 - P(G))}$$

$$P(L_n|Action_n) = \ P(L_{n-1}|Action_n) + \left((1 - P(L_{n-1}|Action_n)\right) * P(T))$$

Figure 1. The equations used to infer student latent knowledge from performance in Bayesian Knowledge Tracing.

Abductive reasoning skills are an important part of the undergraduate learning experience, not just in genetics, but across domains, because they are essential skills in formulating scientific knowledge, and in applying such knowledge to diagnostic tasks.

BKT, when applied to data from the GCT, was moderately successful at predicting transfer, PFL, and retention test performance (Baker et al., 2011a, 2011b; Baker et al., in press). By the end of the student's use of the tutor, BKT could achieve a correlation of 0.353 to transfer for new students, a correlation of 0.285 to PFL for new students, and a correlation of 0.305 to retention for new students. These levels of agreement were clearly better than no agreement, but still far from perfect. However, one positive for this method is that BKT–based predictions of robust learning were able to achieve close to this level of performance with only a subset of the data (the first 30% in the case of transfer). The performance of the BKT model at predicting transfer, as the student completes increasing amounts of the activity, is shown in Figure 2. In other words, the full degree of predictive power available from this method becomes available when the student has 70% more of the activity to complete. Even when prediction is imperfect, it can still be useful for intervention and automated adaptation if it is available early in the learning process.

Inferring Robust Learning in College Genetics with Meta–cognitive Behaviors

In order to improve upon these models, we next distilled features of the students' interaction with GCT that indicated student behaviors relevant to their meta–cognition. As robust learning involves more complex reasoning about material and conceptual understanding than simply whether the student can obtain the correct answer or not, we analyzed some of

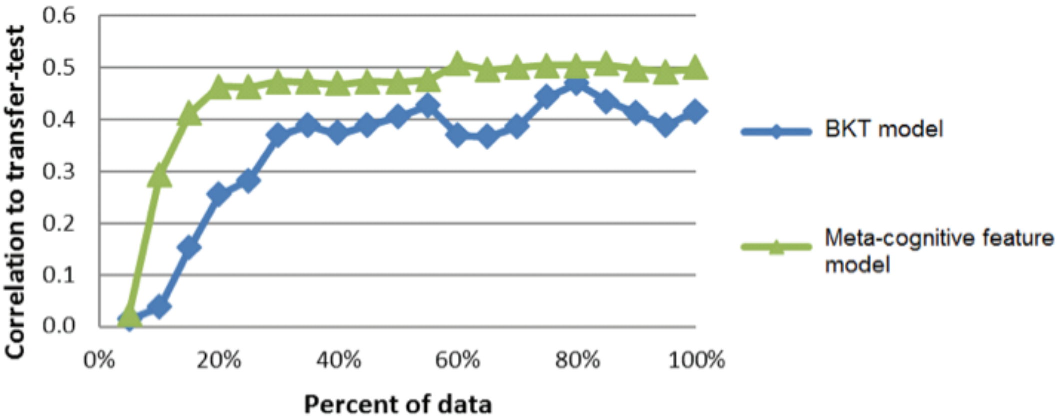


Figure 2. Predicting transfer with first N percent of the data. Graph reproduced with minor modifications from Baker et al. (2011a).

the more complex aspects of student behavior during learning. In doing so, we focused on behaviors that were informative about whether the student was demonstrating meta–cognition, and their engagement with the material. An example of such behavior might be when the software indicates to the student that their response involves a known misconception, and explains why the student's answer was wrong. Does the student pause to think through this explanation, or do they hurry forward without thinking carefully?

The history of research on retention of material, both in research settings and classroom settings, has depended heavily on retesting the same material (or same skill).

A set of 18 features reflective of student thinking were distilled from the students' interactions with the learning system, as shown in Table 1. As also shown in the table, several of these features were found to be individually predictive of PFL and transfer among college students (Baker et al., in press), but only one feature was predictive of retention. When combined into an integrated model (which used some but not all of these features, as some did not provide additional predictive power once other features were incorporated), all three models relied on whether the student sought help when they were struggling, or avoided help. The PFL model also relied upon whether the student paused to self–explain the hints they received. In addition to help seeking, the transfer model relied on whether students made fast actions that did not involve gaming the system (trying to get through the material without learning, for example by systematically guessing; cf. Baker, Corbett, Koedinger, & Wagner, 2004).

This produced the following models of transfer, PFL, and retention:

Transfer = – 1.5613 * HelpAvoidance(1) + 0.2968 * FastNotGaming(7') + 0.8272

PFL = 0.0127 * Spikiness(9) – 0.5499 * HelpAvoidance(1) – 5.3898 * LongPauseAfterHint(4) + 0.8773

Retention = – 2.398 * HelpAvoidance (1) + 0.852

When applied to new students, the transfer model achieved a correlation of 0.396 (Baker et al., in press), the PFL model achieved a correlation of 0.454 (Baker et al., in press), and the retention model achieved a correlation of 0.410. As such, model performance was better than using BKT estimates of student knowledge alone, although only moderately so. By contrast, the models of retention based on these features did not improve on the knowledge–based models.

In addition, these predictions of robust learning were able to achieve nearly this level of performance with only a subset of the data (the first 20% in the case of transfer), moderately faster than the knowledge–based models. In other words, the full degree of predictive power available from this method becomes available when the student has 80% more of the activity to complete, giving plenty of time for interventions designed to improve the robustness of learning. The performance of the meta–cognitive model at predicting transfer, as the student completes increasing amounts of the activity, is shown in Figure 2.

...we focused on behaviors that were informative about whether the student was demonstrating meta–cognition, and their engagement with the material. An example of such behavior might be when the software indicates to the student that their response involves a known misconception, and explains why the student's answer was wrong. Does the student pause to think through this explanation, or do they hurry forward without thinking carefully?

Table 1
Features Distilled for Development of Automated Detectors of Robust Learning

1	Help avoidance: the failure to request help on a skill the student does not know, on their first attempt (Aleven et al., 2006) *%&
1'	Not requesting help on skills the student already knows
2	An extended pause after receiving feedback for a known misconception *
2'	A short pause after receiving feedback for a known misconception *
3	An extended pause after receiving on-demand help messages *%
3'	A short pause after receiving on-demand help messages %
4	An extended pause after receiving on-demand help messages and getting the answer right (Shih, Koedinger, & Scheines, 2008) *%
4'	A short pause after receiving on-demand help messages and getting the answer right %
5	Long pauses on skills that the student probably knows *%
5'	Short pauses on skills that the student probably knows %
6	Off-task behavior, where the student is not working with the system or learning domain for an extended period of time (assessed using automated detector from Baker, 2007)
6'	Long pauses that are not assessed by the detector as off-task
7	Gaming the system, attempting to complete problems without learning the material, for example by systematically guessing or clicking rapidly through hints to get the answer (assessed using automated detector from Baker, Corbett, Roll, & Koedinger, 2008) *%
7'	Fast actions that do not involve gaming the system *%
8	The student's average probability of careless errors, making an error when the student is thought to have obtained the relevant skill (assessed using automated detector from Baker et al., 2010)
8'	The model's average certainty that a careless error is careless, e.g. the average of this construct when the probability is over 0.5
9	The student's average learning per problem step, according to the moment-by-moment learning model (Baker, Goldstein, & Heffernan, 2011) %
9'	The spikiness of the moment-by-moment learning model, e.g. the ratio between the maximum moment-by-moment learning and the average moment-by-moment learning (Baker et al., 2011) %*

Note. Greater operational detail on features is given in (Baker et al., in press). Features predictive of PFL are marked with a *. Features predictive of transfer are marked with a %. Features predictive of retention are marked with a &.

It is useful to know that these measures of meta–cognitive skill are predictive of robust learning in the domain of genetics. However, these measures are potentially applicable at greater scale than simply a single domain. For instance, the help seeking, help avoidance, and self–explanation models used in this analysis were originally developed in the context of mathematics (e.g., Aleven et al., 2006; Shih et al., 2008). In these previous papers, these same three models were shown to correlate to student learning outcomes. As the exact same models can predict learning outcomes both in high school mathematics and in college genetics, our current results – in combination with the previous results published by other authors – suggest that these models may capture aspects of learning skill that are domain–general. An important next step would be to see if these models' predictions are accurate, for the same student, in new domains. Showing that a model predicts learning outcomes in two domains is different than showing that a student's skill is domain general. In one example of this type of research, Sao Pedro and colleagues (2014) found that students who demonstrate scientific inquiry skill in one science domain are likely to be able to demonstrate the same skill in another domain.

Inferring Robust Learning in College Genetics with Moment–by–Moment Learning Models

A third method for inferring robust learning in college genetics that was tried is moment–by–moment learning models. The moment–by–moment learning model (Baker et al., 2011) is a distillate of Bayesian Knowledge Tracing that tries to infer not just the probability that a student has learned a skill by a certain point in a learning activity, but how much they learned at that stage of the activity. This inference is made using a combination of their current estimated knowledge, their behavior during the current learning opportunity, and their performance in the learning opportunities immediate afterwards.

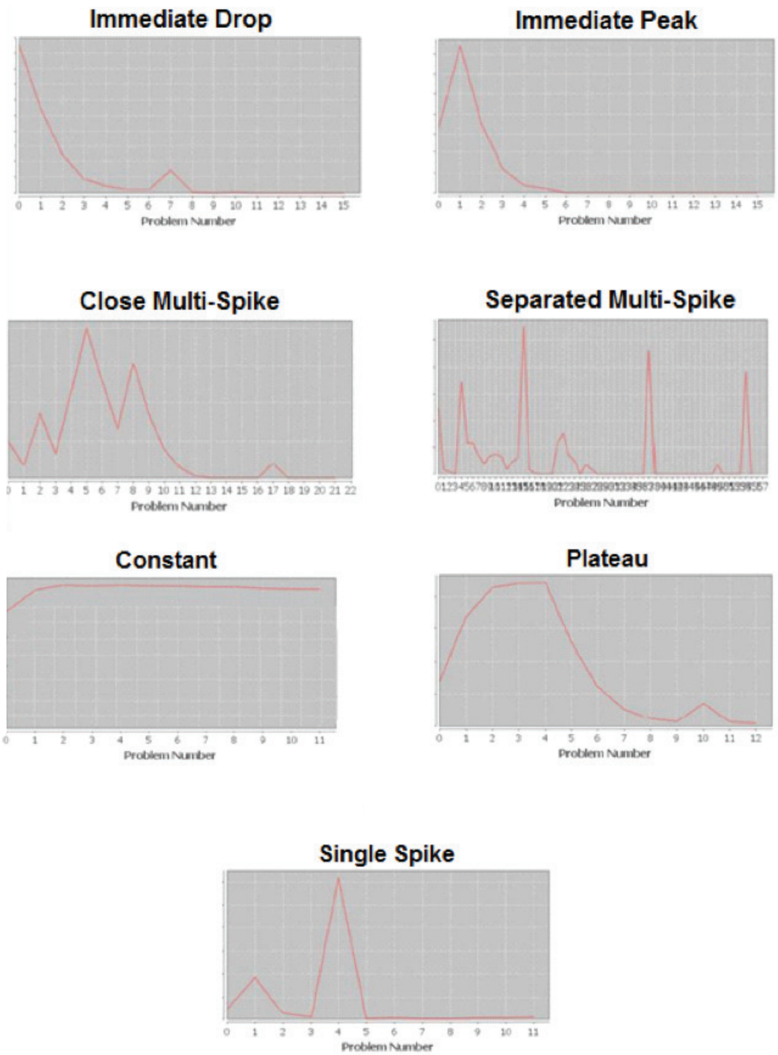


Figure 3. Examples of the visual features of moment–by–moment learning graphs studied by data coders. The x–axis on these graphs represents the number of problems or problem steps where the student has encoundered a specific skill; the y–axis represents the amount of learning inferred to have occurred during the problem step, relative to other problem steps. Note that these graphs show relative differences in learning rather than absolute amounts of learning, in order to facilitate visual interpretation by coders. Graphs reproduced from Baker et al. (2013).

In other words, the full degree of predictive power available from this method becomes available when the student has 80% more of the activity to complete, giving plenty of time for interventions designed to improve the robustness of learning.

The full mathematical details of this model are outside the scope of this paper and take up multiple pages, but are given in full in Baker et al.'s (2011) work. In brief, a combination of the probability of knowledge at the current time (according to BKT) is combined with data on the next two actions, in order to assess the probability of three cases at each time point: The student already knew the skill, the student did not know it but learned it at that time, and the student did not know the skill and did not learn it. Then, machine learning is used to smooth the inferences with additional data on student behavior, including help seeking and pauses. The details of the exact model used to do this smoothing in the case of genetics are given in Baker, Hershkovitz, Rossi, Goldstein, and Gowda's (2013) work.

Visual analysis of moment–by–moment learning over time indicated that there can be very different patterns in different students' learning, or in the learning of the same student for different skills (Baker et al., 2013). Examples are shown in Figure 3. One intuition was that certain patterns during the process of learning may indicate more or less robust learning. This intuition was supported by analyses where human coders labeled graphs by hand in terms of specific patterns, such as plateaus, hillsides, or single–spike graphs, and then these patterns were correlated to robust learning outcomes in GCT (Baker et al., 2013). Examples of these graphs are shown in Figure 3. Some patterns such as plateaus appeared to be correlated to less robust learning, whereas other patterns such as hillsides, where the student learns the skill quickly upon beginning to use the system, appeared to be correlated to more robust learning. These patterns generally held across all three forms of robust learning.

Next, attempts were made to automate this process, distilling mathematical features of the graphs of learning over time, and building these into models to predict robust learning automatically within GCT (Hershkovitz, Baker, Gowda, & Corbett, 2013). The best model of PFL involved the area under the graph (an indicator of total learning), the height of the third–largest peak (the problem step where the third–most learning occurred), and the relative differences both in magnitude and time between the largest peak and the third–largest peak. This model achieved a correlation to PFL of 0.532 for new students, a better performance than the models based on meta–cognitive behaviors or knowledge. This work has not yet been replicated for transfer or retention. However, this model has one disadvantage compared to those models. Although it does not require the application of time consuming post–tests, it cannot infer the robustness of student learning until the student has completed the learning activity, making it less useful for immediate intervention during learning.

Conclusion

In this article, we have discussed multiple ways that robust learning can be inferred within higher education. One popular option is post–tests, whether administered online or on paper. For summative purposes, tests are likely to remain the gold standard option for some time. However, the data from online learning, in combination with educational data mining, provides an alternative with some benefits. Post–tests are time consuming to administer, and cannot be given in real time (particularly for retention tests, which by definition must be administered at a considerable delay). Models that can infer and predict robust learning from learning process data can make predictions which correlate to student robust learning outcomes, predictions which are available to instructors and for personalization within online learning systems much more quickly than paper tests can be available. At some cost to predictive power, predictions can be available as early as when the student has completed only 20% of the learning task. They can also help us to better understand the processes which lead to robust learning.

In our work with the Genetics Cognitive Tutor, we have developed three approaches to inferring robust learning: knowledge–based modeling, metacognitive–behavior–based modeling, and moment–by–moment–learning–based modeling.

In our work with the Genetics Cognitive Tutor, we have developed three approaches to inferring robust learning: knowledge–based modeling, metacognitive–behavior–based modeling, and moment–by–moment–learning–based modeling. The knowledge–based modeling approach was simplest to create as it depended solely on a standard model for measuring learning in online problem–solving; its performance was, however, the weakest. The approach based on modeling metacognitive behaviors required more effort to create; it reached asymptotic performance at inferring transfer and PFL after the student had completed 20% of the learning activity. Finally, the approach based on the moment–by–moment–learning–model was best at inferring PFL, but is not applicable until the student has completed the learning activity.

As such, models like the meta–cognitive behavior model are probably most relevant for use in automated interventions that attempt to infer which students are at risk of developing shallow learning and intervene in real time to enhance their learning. By contrast, models like the moment–by–moment–learning model are probably most relevant for informing instructors after an activity in which students have not developed robust learning, or for recommending additional alternate activities after a student completes an activity without achieving robust learning. Either approach is more work during development than simply creating a test; but these approaches have the potential to speed up assessment and facilitate giving students more rapid learning support.

Beyond their ability to predict tests of robust learning in a specific domain, these types of new measures may point the way to new domain–general assessment of student skills. In particular, the types of help seeking skills used in the meta–cognitive model have the potential to be domain–general, as science inquiry skills have been shown to be (e.g., Sao Pedro et al., 2014). It is not yet clear whether the moment–by–moment learning model indicators of robust learning will also prove general, but this is a valuable potential area for future work.

The importance of robust learning for higher education is clear. The goal of an undergraduate education is not simply to produce mastery of a known set of skills, or awareness of a known set of knowledge, but to prepare students for their future careers, where they will have to be able to transfer their knowledge to new situations and contexts, and where they will need to be prepared for future learning, both in the domains they have studied and in the new areas that will emerge after they complete their studies.

As such, it is important to assess robust learning in higher education, and to support students in developing it. The approaches presented here represent a variety of ways that may make assessment of robust learning more feasible in the higher education context.

AUTHOR'S NOTE

We would like to thank Jaclyn Ocumpaugh for assistance in the early stages of conceptualizing and preparing this article, Annie Levy for assistance in the technical aspects of paper preparation, and the anonymous reviewers for helpful comments and suggestions. We would like to acknowledge Award # DRL–091018 from the National Science Foundation.

References

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). *Toward meta–cognitive tutoring: A model of help seeking with a Cognitive Tutor. International Journal of Artificial Intelligence and Education*, *16*, 101–128.

Bahrick, H. P., Bahrick, L. E., Bahrick, A. S., & Bahrick, P. E. (1993). Maintenance of foreign language vocabulary and the spacing effect. *Psychological Science, 4*, 316–321.

Baker, R. S. J. d. (2007). Modeling and understanding students' off–task behavior in intelligent tutoring systems. *Proceedings of ACM CHI 2007: Computer–Human Interaction*, 1059–1068.

Baker, R. S. J. d. (2013). Learning, schooling, and data Analytics. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on innovations in learning for states, districts, and schools* (pp.179–190). Philadelphia, PA: Center on Innovations in Learning.

Baker, R. S. J. d., Corbett, A. T., & Gowda, S. M. (in press). Generalizing automated detection of the robustness of student learning in an intelligent tutor for genetics. *Journal of Educational Psychology*.

Baker, R. S. J. d., Corbett, A. T., Gowda, S. M., Wagner, A. Z., MacLaren, B. M., Kauffman, L. R., Mitchell, A. P. & Giguere, S. (2010). Contextual slip and prediction of student performance after use of an intelligent tutor. *Proceedings of the 18th Annual Conference on User Modeling,* an intelligent tutor. *Adaptation, and Personalization*, 52–63.

Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off–task behavior in the cognitive tutor classroom: When students "game the system". *Proceedings of ACM CHI 2004: Computer–Human Interaction*, 383–390.

Baker, R. S. J. d., Corbett, A. T., Roll, I., & Koedinger, K. R. (2008). Developing a generalizable detector of when students game the system. *User Modeling and User–Adapted Interaction, 18*(3), 287–314.

Baker, R. S. J. d., Goldstein, A. B., & Heffernan, N. T. (2011). Detecting learning moment–by–moment. *International Journal of Artificial Intelligence in Education, 21*(1–2), 5–25.

Baker, R. S. J. d., Gowda, S., & Corbett, A. T. (2011a). Towards predicting future transfer of learning. *Proceedings of 15th International Conference on Artificial Intelligence in Education*, 23–30.

Baker, R. S. J. d., Gowda, S. M., & Corbett, A. T. (2011b). Automatically detecting a student's preparation for future learning: Help use is key. *Proceedings of the 4th International Conference on Educational Data Mining*, 179–188.

Baker, R. S. J. d., Hershkovitz, A., Rossi, L. M., Goldstein, A. B., & Gowda, S. M. (2013). Predicting robust learning with the visual form of the moment–by–moment learning curve. *Journal of the Learning Sciences, 22*(4), 639–666.

Baker, R., & Siemens, G. (in press). Educational data mining and learning analytics. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences (2nd ed.).*

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.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long–term learning. *Assessment & Evaluation in Higher Education, 31*(4), 399–413.

Bransford, J. D., & Schwartz, D. L. (1999). Rethinking transfer: A simple proposal with multiple implications. In A Iran–Nejad & P. D. Pearson (Eds.), *Review of research in education* (Vol. 22, pp. 61–100). Washington, DC: American Educational Research Association.

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 & Knowledge*. New York, NY: ACM.

Chin, D. B., Dohmen, L. M., Cheng, B. H., Oppezzo, M. A., Chase, C. C., & Schwartz, D. L. (2010). Preparing students for future learning with teachable agents. *Educational Technology Research and Development, 58*(6), 649–669.

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

Corbett, A., Kauffman, L., MacLaren, B., Wagner, A., & Jones, E. (2010). A cognitive tutor for genetics problem solving: Learning gains and student modeling. *Journal of Educational Computing Research, 42*(2), 219–239.

Corbett, A., MacLaren, B., Wagner, A., Kauffman, L., Mitchell, A., & Baker, R. (2013a). Enhancing robust learning through problem solving in the genetics cognitive tutor. Poster paper. *Proceedings of the Annual Meeting* *of the Cognitive Science Society*, 2094–2099.

Corbett, A., MacLaren, B., Wagner, A., Kauffman, L., Mitchell, A., Baker, R. S. J. d. (2013b). Differential impact of learning activities designed to support robust learning in the genetics cognitive tutor. *Proceedings of the 16th International Conference on Artificial Intelligence and Education*, 319–328.

Feng, M., Heffernan, N. T., & Koedinger, K. R. (2009). Addressing the assessment challenge in an online system that tutors as it assesses. *User Modeling and User–Adapted Interaction: The Journal of Personalization Research, 19*(3), 243–266.

Fong, G. T., & Nisbett, R. E. (1991). Immediate and delayed transfer of training effects in statistical reasoning. *Journal of Experimental Psychology: General, 120*, 34–45.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1–28.

Hanna, D. E. (1998). Higher education in an era of digital competition: Emerging organizational models. *Journal of Asynchronous Learning Networks*. http://www.aln.org/alnweb/journal/vol2_issue1/hanna.htm

Hershkovitz, A., Baker, R. S. J. d., Gowda, S. M., & Corbett, A. T. (2013). Predicting future learning better using quantitative analysis of moment–by–moment learning. *Proceedings of the 6th International Conference on Educational Data Mining*, 74–81.

Jastrzembski, T. S., Gluck, K. A., & Gunzelmann, G. (2006). Knowledge tracing and prediction of future trainee performance. *Proceedings of the 2006 Interservice/Industry Training, Simulation, and Education Conference*, 1498–1508.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*, 258–272.

Kinnebrew, J. S., & Biswas, G. (2012). Identifying learning behaviors by contextualizing differential sequence mining with action features and performance evolution. *Proceedings of the 5th International Conference on Educational Data Mining*. Chania, Greece.

Knapper, C. K., & Cropley, A. J. (2000). *Lifelong learning in higher education* (3rd ed.). London, UK: Kogan Page.

Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 61–78). Cambridge, UK: Cambridge University Press.

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

Linn, M., & Chiu, J. (2011). Combining learning and assessment to improve science education. *Research & Practice in Assessment, 6*(2), 5–14.

Mislevy, R. J., Almond, R. G., & Lukas, J. (2004). *A brief introduction to evidence–centered design* (CSE Technical Report 632). The National Center for Research on Evaluation, Standards, Student Testing (CRESST).

Ocumpaugh, J., Baker, R. S., Kamarainen, A. M., & Metcalf, S. J. (2014). Modifying field observation methods on the fly: Metanarrative and disgust in an environmental MUVE. *Proceedings of PALE 2013: The 4th International Workshop on Personalization Approaches in Learning Environments*, 49–54.

O'Reilly, U. M., & Veeramachaneni, K. (2014). Technology for mining the big data of MOOCs. *Research & Practice in Assessment, 9*(2), 29-37.

Pardos, Z. A., Baker, R. S. J. d., Gowda, S. M., & Heffernan, N.T. (2011). The sum is greater than the parts: Enabling models of student knowledge in educational software. *SIGKDD Explorations, 13*(2), 37–44.

Puncochar, J., & Klett, M. (2013) A model for outcomes assessment of undergraduate science knowledge and inquiry processes. *Research & Practice in Assessment, 8*(2), 42–54.

Qiu, Y., Qi, Y., Lu, H., Pardos, Z. A., & Heffernan, N. T. (2011). Does time matter? Modeling the effect of time with Bayesian knowledge tracing. *Proceedings of the International Conference on Educational Data Mining (EDM)*, 139–148.

Reye, J. (2004) Student modeling based on belief networks. *International Journal of Artificial Intelligence in Education, 14*, 1–33.

Romero C., & Ventura. S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*(1), 135–146.

Sao Pedro, M., Baker, R., & Gobert, J. (2012). Improving construct validity yields better models of systematic inquiry, even with less information. *Proceedings of the 20th Conference on User Modeling, Adaptation, and Personalization*, 249–260.

Sao Pedro, M. A., Gobert, J. D., & Baker, R. (2014). The impacts of automatic scaffolding on students' acquisition of data collection inquiry skills. *Roundtable presentation at American Educational Research Association*.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207–217.

Schwartz, D. L., & Martin, T. (2004). Inventing to prepare for future learning: The hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*(2), 129–184.

Shih, B., Koedinger, K. R., & Scheines, R. (2008). A response time model for bottom–out hints as worked examples. *Proceedings of 1st International Conference on Educational Data Mining*, 117–126.

Shute, V. J., & Ventura, M. (2013). *Stealth assessment: Measuring and supporting learning in video games*. Cambridge, MA: MIT Press.

Singley, M. K., & Anderson, J. R. (1989). *Transfer of cognitive skill*. Cambridge, MA: Harvard University Press.

Smith, M. U. (1988, April). *Toward a unified theory of problem solving: a view from biology*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

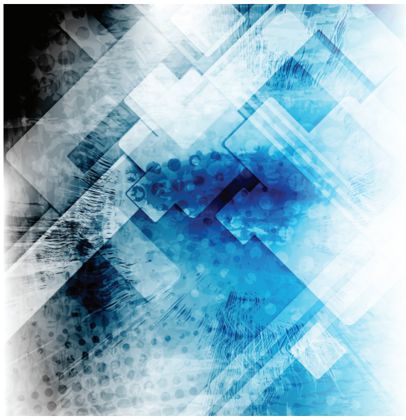
Sperling, J., & Tucker, R.W. (1997). *For–profit higher education: Developing a world–class workforce.* New Brunswick, NJ: Transaction Publishers.

Surber, J. R., & Anderson, R. C. (1975). Delay–retention effect in natural classroom settings. *Journal of Educational Psychology, 67*(2), 170–173.

VanLehn, K., Jones, R., & Chi, M. T. H. (1992). A model of the self–explanation effect. *Journal of the Learning Sciences, 2*(1), 1–59.

Vardi, M. Y. (2012). Will MOOCs destroy academia? *Communications of the ACM, 55*(11), 5.

Wang, Y., & Heffernan, N. (2011). Towards modeling forgetting and relearning in ITS: Preliminary analysis of ARRS data. *Proceedings of the 4th International Conference on Educational Data Mining*, 351–352.



Abstract

Current trends and challenges in higher education (HE) require a reorientation towards openness, technology use and active student participation. In this article we will introduce Social Learning Analytics (SLA) as instrumental in formative assessment practices, aimed at supporting and strengthening students as active learners in increasingly open and social learning environments. The analysis of digital traces of students' learning behaviors provides insight into learning opportunities and can raise students' awareness about where to be and whom to join. Against the background of these HE trends and challenges, we discuss opportunities for applying SLA to support open learning practices, that will move students from awareness to productive engagement in learning activities that promote co–construction of knowledge.

AUTHORS

Maarten de Laat, Ph.D.
Open University of the Netherlands

Fleur R. Prinsen, Ph.D.
Open University of the Netherlands

Social Learning Analytics: Navigating the Changing Settings of Higher Education

Higher education (HE) is increasingly seen as needing to change in ways that meet the transformation of our times (Warner, 2006). For HE institutions to remain relevant to the social settings in which they exist, Wiley and Hilton III (2009) argue that creating an institutional culture of openness is the most pressing priority. Massive Open Online Courses (MOOC) development and Open Educational Resources (OER) are demonstrative of the societal movement towards more openness.

Several developments towards more openness are already emerging. Institutions are becoming transparent and are starting to promote open communication and open scholarship (Czerniewicz, 2013). Changing expectations and the adoption of progressive technology challenge HE to replace its model of delivering education with one that promotes a stronger focus on student participation and collaborative learning, shifting the focus to more active engagement in knowledge co–creation, in an attempt to leave the transmission model of knowledge behind. Pedagogical designs are evolving towards providing open access, promoting networked social activities, and linking education with professional learning communities and lifelong learning to provide their students with broader opportunities to access social capital. This means an increased focus on community learning as well as collaborative, interactive and participatory learning (e.g., Tucker et al., 2013; Zhao & Kuh, 2004).

CORRESPONDENCE

Email
Maarten.deLaat@ou.nl

Some other telling examples of how learning settings are changing are offered by Bayne, Gallagher and Lamb (2014) and Gourlay and Oliver (2013). They explore students' uses and experiences of spaces, as sites of scholarly activity. Bayne et al. argue that HE has taken little account of how space – under the influence of new technologies – is increasingly seen by students as a dynamic entity produced by social practices. Learning spaces have become more fluid, democratic, influenced now by the promises of accessibility to all from the open education movement (see also Knox, 2013), at the same time transforming educational practices (e.g., Ehlers, 2011). The study by Gourlay and Oliver (2013) reveals the complexity of students' orientations towards technology and also the distributed nature of their learning practices across multiple spaces. Thus, learning practices are changing towards increased connectedness, personalization, participation, and openness; the emergence and popularity of MOOCs as new spaces for learning can be seen as an illustration of this (Macfadyen, Dawson, Pardo, & Gasević , 2014).

We are left, however, with an important question: How do we assess and facilitate productive social connectivity and mobility in these open learning spaces? When learning is designed around social engagement and interaction, there is a need to develop new ways of understanding and assessing student social mobility. We need to be able to promote and monitor student engagement and offer them direct ways to reflect on their learning activities – and that of others – raising awareness about the opportunities these open learning practices have to offer. In this article we explore what a newly developing design discipline (Knight, Buckingham Shum, & Littleton, 2014), called learning analytics, can contribute to address this.

Below we will introduce Social Learning Analytics (SLA) as an instrument in formative assessment practices aimed at supporting and strengthening students as active learners in the process of becoming practitioners. SLA, applied in open HE settings, will help students make informed decisions about where to be and whom to join for their learning , by tracking and visualizing indicators of social learning behaviors and patterns in those behaviors. This will raise awareness and equip students with the kind of orientations necessary to meet the demands of the emerging open networked society.

Trends and Challenges in Higher Education

The changes that HE is facing have recently been substantiated by the NMC Horizon Report > Higher Education Edition (Johnson et al., 2013). This report identifies key trends that influence the HE future agenda, covering use of technology, change in student participation and challenging models for teaching and learning.

Developments in technology use and availability have been a strong driver for change in behavior and learning. The growing ubiquity of social media and an ongoing integration of online, hybrid and collaborative learning are identified trends that already have impacted HE and we have witnessed or are witnessing the effects of it. Social media has opened the traditional organizational boundaries of HE institutions and is changing scholarly communication enabling less formal "two way dialogues between students, prospective students, educators, and the institution" (Johnson et al., 2013, p. 8). Increased social media use transforms HE from institutionalized into more open scholarly practices, with knowledge and content becoming increasingly open and accessible (Czerniewicz, 2013). At the same time, hybrid or blended forms of teaching and learning offer more freedom in interactions with and between students, and encourage collaboration, thus reinforcing real world skills.

Pedagogical designs are evolving towards providing open access, promoting networked social activities, and linking education with professional learning communities and lifelong learning to provide their students with broader opportunities to access social capital.

In response to openness, institutions for HE are redesigning physical settings as well, trying to combine the best of both worlds. These modern campuses, also referred to as *sticky campuses* (e.g., Dane, 2014; Lefebvre, 2013), are designed to offer a mixture of formal and informal learning experiences aimed to provide a quality rich environment where students want to be, not only to study, but to socialize and learn. As such these HE learning landscapes are transforming into open learning spaces aimed at becoming a vibrant social hub where people meet and connect 24/7, on and off–line. For example, the University of South Australia recently opened their Jeffrey Smart building on the City West Campus in Adelaide. This building has been designed to be a lively learning hub and open space used by students, staff and professionals. The open space has been developed for students to come and interact with their peers, build networks and communities, facilitate collaborative learning, share experiences and knowledge to enhance and enrich their university learning experience. Engaging in open practices, and the ability to build and utilize rich social networks are essential skills and capabilities students require to be proficient learners in an increasingly networked society.

Inspired to some extent by the technological possibilities, some of the traditional roles in HE teaching and learning practices are changing as well. Education becomes more personalized and students are becoming active participants emphasizing learning by making and creating instead of passively consuming content. Some HE campuses are building living labs to promote a holistic approach to teaching or are using real built environments for user–centered research and the creation of a collaborative learning platforms (e.g., Masseck, 2013). Through advanced engagement in hybrid learning environments, students also leave an increasingly clear trail of analytics data that can be mined for insights. Utilizing student data for learning analytics in itself has become a new trend, and "there is a growing interest in developing tools and algorithms for revealing patterns inherent in those data and then applying them to the improvement of instructional systems" (Masseck, 2013, p. 12).

Finally another trend is that HE institutions are looking to provide a more diverse offering of opportunities and access to quality education. MOOCs, for instance, are:

> Enabling students to supplement their education and experiences at brick–and–mortar institutions with increasingly rich, and often free, online offerings. Downes and Siemens envisioned MOOCs as ecosystems of connectivism – a pedagogy in which knowledge is not a destination but an ongoing activity, fueled by the relationships people build and the deep discussions catalyzed within the MOOC. That model emphasizes knowledge production over consumption, and new knowledge that emerges from the process helps to sustain and evolve the MOOC environment. (Johnson et al., 2013, p. 26)

Social Learning: Participation, Co–Creation and Becoming

The above trends have among else in common that they challenge HE institutions to embrace social theories of learning. Learning is increasingly seen to be most effective when it is collaborative and social in nature (De Laat, 2012; Siemens, 2005). In social forms of learning, the focus is on the co–construction of knowledge, meaning and understanding. This takes into consideration how the practical, social (learning) situation influences individual and collective outcomes of learning. Learning in a social context is a process of meaning–making, where this meaning can be based upon prior experiences as well as the more immediate social context in which something is learned. Meaning is made through negotiation among the various actors participating in a learning context.

When learning is designed around social engagement and interaction, there is a need to develop new ways of understanding and assessing student social mobility.

New metaphors describing social learning have gained currency and are used to develop a language for learning that emphasizes important social aspects such as participation, co–construction and becoming (Hager & Hodkinson, 2009; Packer & Goicoechea, 2000). In this context the application of 21st century skills such as collaboration (working in teams, learning from and contributing to learning of others, social networking skills, empathy in working with diverse others), creativity and imagination (economic and social entrepreneurialism, considering novel ideas and leadership for action) is emphasized (see Dede, 2010 for an overview).

Whereas the 21st century skills focus mostly on participation and co–construction, the notion of learning as becoming (Colley, James, Diment, & Tedder, 2003; Hodkinson, Biesta, & James, 2008) has been explored for example by Shaffer (2004). He provides inspiring examples, in which students' identity development is stimulated through the adoption of practices associated with the ways of knowing of particular professional communities. Shaffer developed extended role playing games, simulating professional learning. Professions have their own ways of knowing, of deciding what is worth knowing and of adding to the collective body of knowledge and understanding of a community. Shaffer's studies show that students can incorporate these elements into their identities when engaged in games. One epistemic game Shaffer writes about is SodaConstructor, tapping into the ways of knowing of engineering and physicists' communities. In the game participants can design their own virtual creature, applying (and thereby showing understanding of) fundamental concepts from physics and engineering. They test their ideas through a simulation of how this creature would operate once gravity, friction and muscles enter the equation. This way they can mimic the creative thinking of engineers: creating designs, building them, and then testing alternatives as well.

HE students, seen through the new metaphorical lenses of participation, co–creation and becoming, are thus learning to engage in open educational practices. Open educational practices are implemented through open pedagogies (Ehlers, 2011). There are gradations in how open these pedagogies are (see Figure 1), depending on how much freedom students have to develop open practices and the degree of involvement of others in their learning.

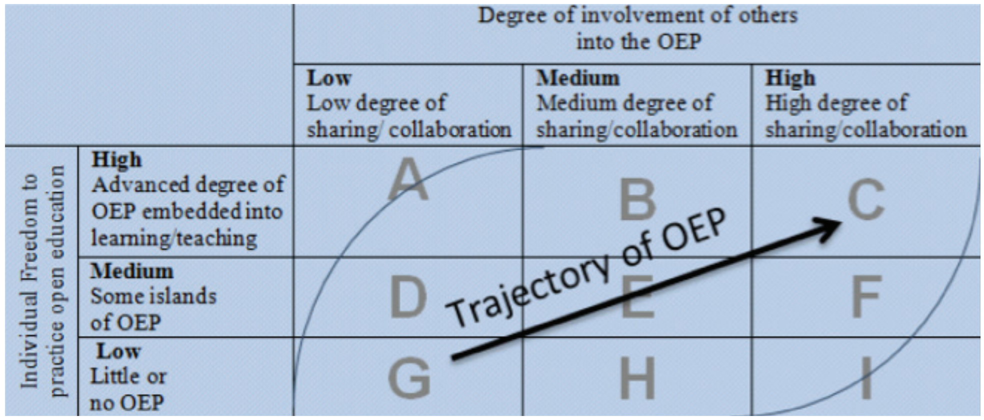


Figure 1. Diffusion of open educational practice (from Ehlers, 2011).

New forms of assessment also ensue from these changing perspectives on learning; monitoring and openly valuing student engagement and helping students become more aware and able to reflect on productive social learning practices. Social learning analytics are instrumental in this.

Social Learning Analytics

With the new trends in HE come another trend, giving rise to data–driven learning and assessment and paving the way for learning analytics (LA). Some institutions – like Purdue University and Marist College – are forerunners who actively implement LA tools to help manage learning and organizational strategies. Other organizations are still observing these developments, but they are increasingly aware that a data–driven understanding of learning and assessment is an approach they need to embrace. It is evident that LA is an emerging field that, like other areas where analytics is applied, (e.g., HE marketing and management), is drawn to massive computerized activity and big data with the means to improve and support learning. LA concerns 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 (Siemens, 2013).

Through advanced engagement in hybrid learning environments, students also leave an increasingly clear trail of analytics data that can be mined for insights.

A particular area within LA capitalizes on institutional big data used to track and evaluate student behavioral patterns. Learning Management Systems, for instance, enable the collection of data on student demographics, measures of (prior) academic performance and student behavior. These aspects of LA are more concentrated on the management of learning and understanding personal (background) characteristics, whereas another research area concentrates on harnessing data to understand student connectivity and the development of social relationships, and how this can be used to promote learning through social interaction. This work, referred to as social learning analytics (SLA; Buckingham Shum & Ferguson, 2012), is aimed at analyzing ongoing learning and group dynamic processes, course design features and resulting outcomes in terms of collaborative practice, development of learning communities, in formal or informal settings, design and development of social learning systems that utilize networked connectivity and learning partnerships (Haythornthwaite, De Laat, & Dawson, 2013).

Buckingham Shum and Ferguson (2012) make a useful distinction between inherently social analytics, and socialized analytics. Inherently social analytics only make sense in a collective context. Socialized analytics are relevant as personal analytics, but can also be usefully applied in social settings (e.g., disposition analytics; intrinsic motivation to learn lies at the heart of engaged learning and innovation). An important example of an inherently social analytic, as discussed by Buckingham Shum and Ferguson, is social network analysis. Social network analysis can be used to investigate networked learning processes through analysis of the properties of connections, the roles people take in their learning relations and the significance of certain network formations. It can aid in understanding how people develop and maintain relations to support learning (Haythornthwaite & De Laat, 2010).

Although there are some SLA tools available to support micro level social learning, such as support for collaborative learning processes in small groups and community learning, what is largely missing are SLA tools that build on large scale social mobility and help students to become more aware of productive social connectivity. Social awareness about meaningful networked activity on this meso or even macro level within, across and beyond HE institutions (in relation to the trends discussed earlier) is needed to support productive social learning associated with the living social hubs that HE institutions aspire to be (e.g., Hemmi, Bayne, & Land, 2009). Through social learning analytics, based on data about student movements, we might be able to provide a better insight in the social dynamics and networked learning opportunities that these HE social hubs and sticky campuses have to offer. It allows students to become aware of relevant social mobility, important (community) events and networked activity that suits their needs as a learner and helps them to make informed choices about where and when to participate.

Learning is increasingly seen to be most effective when it is collaborative and social in nature.

Below we discuss a model (see Figure 2) that focuses on what we call social enterprise analytics in an attempt to address these social mobility challenges and we will present a few examples of what such SLA tools might look like. This model is a combination of raising awareness about social learning activity as well as leveraging a culture of knowledge and value creation. We think it is important to not only develop tools but pay attention to the context in which these learning practices take place. We need to pay more attention to the social and cultural aspects that characterize learning, rather than keeping our focus mainly on learning outcomes and products (De Laat, 2012). This will require HE institutes to review their approach to learning and try to move from a results driven culture towards a culture that embraces the value of being engaged in social learning processes. This calls for rewarding engagement in practices where students are connected in networks and communities, and understand and assess how they create value.

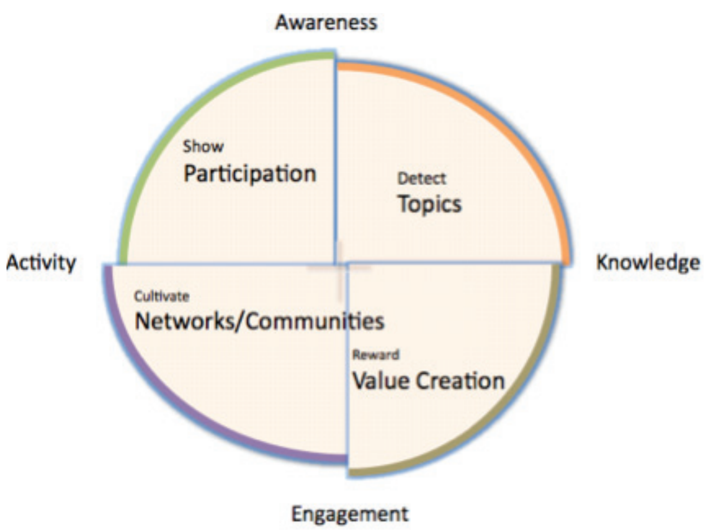


Figure 2. Social enterprise analytics (De Laat, 2014).

Analytics can provide the tools that help detect and visualize real time activity patterns of people (students, staff and professionals) and their knowledge. On the one hand these analytics can help to take the pulse of HE organizations and reveal people's learning activity and movement; this way, learners can find out what is currently going on and who are the main drivers of these activities. Finding ways to identify, access, and assess informal emerging activity and topics will be a way to connect people to learning and make informed decisions about participation and develop learning friendships. The top half of the model is therefore aimed at increased awareness in order to link people to content (and vice versa), whereas the lower part is concerned with leveraging a culture of knowledge. Here the focus is on cultivating networks and communities and promote student autonomy and increased responsibility. More openness means less control and planning by the formal educational curriculum and increases student flexibility and freedom to regulate their learning informally and engage in (professional) networks that contribute to their learning goals. For this, one might stimulate student engagement by joining associated, active networks and communities in with their courses and optimize students learning and develop new ways to appreciate and reward value creation (Wenger, Trayner, & De Laat, 2011).

Challenges for Social Learning Analytics

As a relatively new field, SLAs have their own challenges to overcome. A critique often voiced about LA in general is its atheoretical nature. It is often incorrectly assumed that data speak for themselves, but it is important to consider that LA and pedagogy are both bound up in beliefs about what knowledge is. "The ways that we assess, the sorts of tasks we set and the kinds of learning we believe to take place (and aim for) are bound up in our notions of epistemology" (Knight, Buckingham Shum, & Littleton, 2014, p. 77). Assessment instruments come with assumptions about the nature of knowledge and how it comes about. For instance, when knowledge is understood as being distributed and co–constructed among actors in a network of practice, student success is reframed as being well–connected to the learning resources within a specific network. Different approaches have different analytic implications (for other examples see Knight et al., 2014), which means analytics can suffer from interpretative flexibility (Hamilton & Feenberg, 2005) when not properly embedded in a theoretical framework.

In the game participants can design their own virtual creature, applying (and thereby showing understanding of) fundamental concepts from physics and engineering. They test their ideas through a simulation of how this creature would operate once gravity, friction and muscles enter the equation. This way they can mimic the creative thinking of engineers: creating designs, building them, and then testing alternatives as well.

There are also some challenges related to data collection methods. Not all relevant learning traces can be captured digitally and some indicators are not very reliable; e.g., if a student prints out a resource instead of reading it online, the reading time is not a reliable indicator for how much the student has learned, and having a browser window open does not necessarily mean students are reading either. These problems will either have to be treated as measurement errors, or might in the future be addressed by additional tools, e.g., by applying eye–tracking.

Finally, the use of SLA may sometimes raise ethical issues, which need not be overlooked (Slade & Prinsloo, 2013). With LA becoming part and parcel of educational practice, students should take part in shaping and possibly reshaping this new practice of learning; the use of LA should be transparent to them. In addition, Slade and Prinsloo (2013) point out that student success is a multidimensional phenomenon and rather than applying LA in a routine way, LA should function to continuously improve our understanding of how to reach positive outcomes for students (and we would add, with students). We agree with Nissenbaum (2009) that students have a right to an appropriate flow of personal information. Nissenbaum suggests the concept of contextual integrity for LA, where what is considered appropriate will vary from context to context (depending on local "immediately canonical activities, roles, relationships, power structures, norms (or rules), and internal values (goals, ends, purposes)" (p. 132). For instance, as students engage with online activities (e.g., in a Learning Management System), data are generated as a by-product of this activity, including patterns of questions posed and answered (Buckingham Shum & Ferguson, 2012). Frequently student involvement is mandatory in this context, but participation thereby should not be too easily considered a measure of learning outcome. When LMS's are designed to provide students with a stimulating learning environment and at the same time to effectively manage student engagement, these are the values internal to this LMS (its goals, ends, purposes) and these should be apparent.

Contemporary Examples

Through SLA, productive social learning processes and arrangements can be identified and made visible, so that they can be assessed and actions can be taken on them. In this section we highlight some contemporary examples of SLA tools and practices we are working on.

Increase Awareness and Participation

NetMap (De Laat, Dawson, & Bakharia, 2014) is prototype software developed at the University of South Australia in collaboration with the Open University of the Netherlands to provide a medium for students to unlock the potential of previously hidden informal learning networks. The software centers on facilitating the development of collaborative student interactions. As such, NetMap serves as a kind of dating system for developing learning relationships in the physical space using GPS location data combined with information about the topics that people are working on. The central idea is to map informal networks and raise the awareness of potential learning ties for situated learning. When one enters the space they can use the software to select the topics they are interested in, browse people's profiles and find out where they are located in the open space as their current GPS position is highlighted on the map. Based on this information one will be able to quickly find peers who are open to sharing and collaboration on this particular topic. NetMap will additionally be used by tutors, university support services, or faculty and could be taken up by industry to open up more informal student engagements and promote stronger connections into specific industry groups.

Increase Awareness and Cultivate Networks

In order to find relevant and up–to–date information, students and teachers in their learning activities are turning to online resources more than ever before. Google Scholar is a popular example, but students can also access online professional communities for the materials they are looking for. Professionals and students meet each other in open practices where they share information and learn from each other. LA can help connect students with content but also with other knowledge workers to connect to. Students, like other knowledge workers, face an ever increasing amount of information. Consequently, it is getting increasingly difficult for them to remain aware of relevant content, people, activities, and events. One could claim that all knowledge workers face similar challenges; they generally are connected with several knowledge communities at the same time. The example below illustrates how social analytics can provide support.

Contemporary knowledge workers are in need of tools and techniques that help them to stay on a high awareness level (Reinhardt, 2012) and thus retain productive connections to their networks and the knowledge developments in their domain. Reinhardt, Wilke, Moi, Drachsler and Sloep (2012) showed that awareness of researchers in research networks can be enhanced by tools employing social analytics. They first explored the semantic connections between content and people in research networks by analyzing social media artifacts and scientific publications, visualizing the resulting networks to show how researchers might be more aware of activities and interactions therein. They then designed a widget–based dashboard that was meant to support researchers' awareness in their daily working routine. Their research showed the dashboard was easy to use, was less time consuming than similar technologies, user friendly and raised the level of awareness, helping researchers carry out their tasks more effectively (Reinhardt, Mletzko, Drachsler, & Sloep, 2011). Finally they proposed an event management system to help strengthen the ties between researchers and lead to enhanced awareness of relevant information.

Some institutions, like Purdue University and Marist College, are forerunners who actively implement LA tools to help manage learning and organizational strategies.

Cultivate Networks and Value Creation

Engaging in networked learning means that learners need to be in touch with others to participate in constructive conversations (Haythornwaite & De Laat, 2010). To help stimulate, monitor and evaluate such discussion activities an SLA tool was developed to visualize them in real time (Schreurs, De Laat, Teplovs, & Voogd, 2014). This tool was implemented on a MOOC platform to support Dutch teachers' HE training in assessment. The course was introduced through a live webinar in which discussions were held. Forum discussions were subsequently moderated by experts in the field of assessment, emails were sent out to stimulate participation and more live discussions were planned. The tool helped to visualize the learning relationships between users, based on their contributions to the discussion forums. Since the real pay–offs materialize when stakeholders interact with the analytics, thus rendering their connected world more visible (De Laat & Schreurs, 2013), the design allowed the participants to use the plug–in as a social–learning browser to locate people who are dealing with the same learning topics. They could also identify central people in the network; identify the most active ones as well as identify potential experts. Not only does the tool afford reflection by learners on how to interact with peers for learning purposes, their educators can "use the plug–in to guide students in the development of networked learning competences and can gain insight into the ability of groups of students to learn collectively over time, detect multiple (isolated) networks, connect ideas and foster collaboration beyond existing boundaries" (Schreurs et al., 2014, p. 47).

Conclusion and Discussion

HE institutions aspire to be living social hubs, supporting productive social learning and awareness of meaningful networked activity, across and beyond the institutions themselves. When learning is designed around social engagement and interaction there is a need to develop new ways of understanding and assessing student social mobility. Through SLA, based on data about student connectivity and activity, we might be able to provide a better insight in the social dynamics and networked learning opportunities that these HE social hubs and sticky campuses have to offer; supporting students' awareness of important (community) events and networked activity more closely tailored to their learning needs. This will help them make informed choices about where and when to participate.

Assessment instruments come with assumptions about the nature of knowledge and how it comes about…Different approaches have different analytic implications, which means analytics can suffer from interpretative flexibility when not properly embedded in a theoretical framework.

Reflecting on the trends and challenges that HE is faced with, we propose a model that explicitly pays attention to the social and cultural aspects that characterize learning (participation, co–construction and becoming), calling for the rewarding of engagement in practices, where students are connected in networks and communities, and understand and assess how they create value. This model promotes open and transparent information about social learning activity accessible to all participants. This is based on the conviction that learning analytics tools should enrich people's ability to learn and help them to make informed choices about learning opportunities that are available to them.

References

Bayne, S., Gallagher, M. S., & Lamb, J. (2014). Being 'at' university: The social topologies of distance students. *Higher Education, 67*(5), 569–583.

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

Colley, H., James, D., Diment, K., & Tedder, M. (2003). Learning as becoming in vocational education and training: Class, gender and the role of vocational habitus. *Journal of Vocational Education and Training, 55*(4), 471–498.

Czerniewicz, L. (2013, April). Power and politics in a changing scholarly communication landscape. Paper presented at the 34th Conference of the International Association of Scientific and Technological University Libraries (IATUL), Cape Town.

De Laat, M. (2012). *Enabling professional development networks: How connected are you?*. Open University of the Netherlands. Available athttp://www.ou.nl/documents/14300/1097433/Oratie_de+Laat+_WEB_271112.pdf

De Laat, M. (2014). *Social enterprise analytics*. Retrieved from http://www.open.ou.nl/rslmlt/Social_Enterprise_Analytics_DeLaat.pdf

De Laat, M., Dawson, S., & Bakharia, A. (2014). *NetMap: Visualization of location–based networks using GPS*. Adelaide: University of South Australia.

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.

Dane, J. (2014). What will the campus of the future look like? [web log]. Retrieved 9, 2014, from http://www.woodsbagot.com/news/what–will–the–campus–of–the–future–look–like

Dede, C. (2010). Comparing frameworks for 21st century skills. In J.A. Bellanca & R.S. Brandt, *21st century skills: Rethinking how students learn*. Bloomington, IN: Solution Tree Press.

Ehlers, U. D. (2011). Extending the territory: From open educational resources to open educational practices. *Journal of Open, Flexible and Distance Learning, 15*(2), 1–10.

Gourlay, L., & Oliver, M. (2013). Beyond the 'the social': Digital literacies as sociomaterial practice. In R. Goodfellow, & M. Lea (Eds.), *Literacy in the digital university: Learning as social practice in a digital world*. London, UK: Routledge.

Hager, P., & Hodkinson, P. (2009). Moving beyond the metaphor of transfer of learning. *British Educational Research Journal, 35*(4), 619–638.

Hamilton, E., & Feenberg, A. (2005). The technical codes of online education. *Techné: Research in Philosohy and Technology, 9*(1).

Haythornthwaite, C., & De Laat, M. (2010, May). Social networks and learning networks: Using social network perspectives to understand social learning. In *7th International Conference on Networked Learning*.

Haythornthwaite, C., De Laat, M., & Dawson, S. (2013). Learning analytics. *American Behavioral Scientist*.

Hemmi, A., Bayne, S., & Land, R. (2009). The appropriation and repurposing of social technologies in higher education. *Journal of Computer Assisted Learning, 25*(1), 19–30.

Hodkinson, P., Biesta, G., & James, D. (2008). Understanding learning culturally: Overcoming the dualism between social and individual views of learning. *Vocations and Learning, 1*(1), 27–47.

Johnson, L., Adams Becker S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H. (2013). *The NMC horizon report: 2013 higher education edition*. Austin, TX: The New Media Consortium.

Knight, S., Buckingham Shum, S., & Littleton, K. (in press). Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space. *Journal of Learning Analytics*.

Lefebvre, M. (2013, August). The library, the city, and infinite possibilities–Ryerson University's Student Learning Centre Project. Paper presented at the World Library and Information Congress (IFLA), Singapore.

Macfadyen, L. P., Dawson, S., Pardo, A., & Gasevic, 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.

Masseck, T. (2013). LOW3 – A Living Lab approach for a holistic learning and open innovation platform for sustainability at UPC – Barcelona Tech. Retrieved from http://sites.guninetwork.org/posters2013/217_96.pdf

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.

Packer, M. J., & Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational Psychologist, 35*(4), 227–241.

Reinhardt, W. (2012). Awareness Support for Knowledge Workers in Research Networks (Doctoral dissertation). April, 05, 2012, Open University in the Netherlands (CELSTEC), Heerlen, The Netherlands.

Reinhardt, W., Mletzko, C., Drachsler, H., & Sloep, P. B. (2011). AWESOME: A widget–based dashboard for awareness–support in research networks. In Proceedings of The PLE Conference 2011. July, 11–13, 2011, Southampton, UK.

Reinhardt, W., Wilke, A., Moi, M., Drachsler, H., & Sloep, P. (2012). Mining and visualizing research networks using the Artefact–Actor–Network approach. In *Computational social networks* (pp. 233–267). London, UK: Springer.

Schreurs, B., De Laat, M., Teplovs, C., & Voogd, S. (2014). Social learning analytics applied in a MOOC–environment. *e–Learning Papers, 26*, 45–48.

Shaffer, D. W. (2004). Epistemic frames and islands of expertise: Learning from infusion experiences. In Y. Kafai, W. A. Sandoval, N. Enyedy, A. S. Nixon, & F. Herrera (Eds.), *Proceedings of the Sixth International Conference of the Learning Sciences* (pp. 473–480). Mahwah, NJ: Erlbaum.

Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1), 3–10.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist.* doi: 0002764213498851

Slade, S., & Prinsloo, P. (2013). Learning analytics ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529.

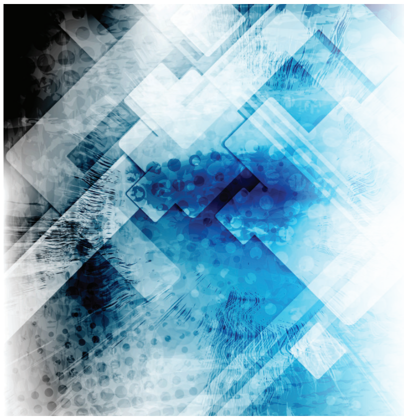
Tucker, B. G., Kazmer, D. O., Bielefeldt, A. R., Paterson, K., Pierrakos, O., Soisson, A., & Swan, C. (2013). Principles of sustaining partnerships between higher education and their larger communities: Perspectives from engineering faculty engaged in learning through service. *International Journal for Service Learning in Engineering, Humanitarian Engineering and Social Entrepreneurship*, 48–63.

Warner, D. (2006). *Schooling for the knowledge era*. Camberwell, UK: ACER Press.

Wenger, E., Trayner, B., & De Laat, M. (2011). Promoting and assessing value creation in communities and networks: A conceptual framework. *The Netherlands: Ruud de Moor Centrum*.

Wiley, D., & Hilton III, J. (2009). Openness, dynamic specialization, and the disaggregated future of higher education. *The International Review of Research in Open and Distance Learning, 10*(5).

Zhao, C. M., & Kuh, G. D. (2004). Adding value: Learning communities and student engagement. *Research in Higher Education, 45*(2), 115–138.



Abstract

This article explores the challenges that students face in navigating the curricular structure of post–secondary degree programs, and how predictive analytics and choice architecture can play a role. It examines Degree Compass, a course recommendation system that successfully pairs current students with the courses that best fit their talents and program of study for upcoming semesters. Data are presented to demonstrate the impact that this system has had on student success. In particular the data will show that by closing the information gap, this system is able to close the educational achievement gap for low–income and minority students.

AUTHOR

Tristan Denley, Ph.D.

Tennessee Board of Regents

How Predictive Analytics and Choice Architecture Can Improve Student Success

It has been a longstanding reality that success in higher education is very uneven across the population of the United States. Consistently over the last three decades racial minority, low–income, and first generation students have earned post–secondary degrees at substantially lower rates than their counterparts. Although the degree–attainment rates for these three groups have increased over that time horizon, those improvements have not kept pace with the degree attainment rates of students in general (NASH & The Educational Trust, 2009; NCES, 2012; U.S. Census Bureau). The most recent IPEDS data show that whilst 49 percent of white students who began college in 2007 graduated with at least an associates degree in 6 years, 37 percent of their African American counterparts, and 33 percent of Hispanic students graduated. While the rate at which low–income students enroll in higher education has doubled since the 1970s the graduation rate for these students has only grown from 7 percent to 10 percent (NASH & The Educational Trust, 2009; Postsecondary Education Opportunity.[1]) First generation students begin to trail their peers as early as their first year, earning 18 credits, on average, compared to the 25 credits earned by students whose parents have degrees (Chen & Carroll, 2005). In fact, similar patterns emerge for minority, low–income, and first generation students in every success metric governing student progress through college when compared with their white, higher–income or non–first generation peers (Kelly, 2005; Lumina Foundation, 2014; NASH & The Educational Trust, 2009).

CORRESPONDENCE

Email

Tristan.Denley@tbr.edu

These attainment gaps appear to be significantly influenced by information gaps. First generation, low–income and minority students often do not have the advice system that surrounds students whose parents or other relatives have been to college. Information is certainly available to these students, but without knowledge of the structure and nomenclature of higher education they are unable to even frame the questions that would enable them to become informed (Diamond et al., 2014; Hagelskamp, Schleifer, & DiStasi, 2013; Kadlec, Immerwahr, & Gupta, 2014).

[1] http://www.postsecondary.org/

The process of navigating institutions from admission to graduation involves large numbers of crucial decisions, and once again, the information gap plays its part in the achievement gap. Despite the advantages to having a clear direction of study (Jenkins & Cho, 2012), one third of first generation students begin college without identifying a major or program of study, whereas only 13 percent of their peers with college–going parents do so (Chen & Carroll, 2005). Students select their majors with little information about what is involved in successfully completing the program, and often discover too late that the picture they had of that discipline is very different from the reality (Kirst & Venezia, 2004; Smith & Wertlieb, 2005). Low–income and minority students express less knowledge of programmatic demands than their peers. Although students may think that they have an interest in a particular area, they receive little information about whether their academic abilities create a realistic chance of successfully completing that program. What is more, they may associate each discipline with a limited number of careers, and often eliminate disciplines from their list of choices because those jobs are unappealing, without realizing the true variety of career opportunities that lie on the other side of graduation.

First generation, low–come and minority students often do not have the advice system that surrounds students whose parents or other relatives have been to college.

As challenging as the factors involved in choosing the right degree program are, navigating a degree program is no less crucial or challenging. Each student must choose from a variety of courses that satisfy the requirements of their general education core, and then their various degree program requirements. Ideally students would make strategic decisions about which courses are most likely to lead to their success. Instead, they are faced with making choices between courses that, ahead of time, they are not in a position to distinguish between. Indeed higher education has been described as a "post–experience good" (Diamond et al., 2014), since not only is it difficult to envisage or evaluate the experience of studying a particular course or program before hand, the true benefits of that study may not be understood until long into the future. Advisors are often well equipped to provide valuable advice in their own field. But, most programs require students to take courses from across the full spectrum of disciplines, and advisors find themselves challenged to offer useful advice in disciplines far from their own. As higher education funding has become more and more depleted, even access to this advice is far from guaranteed (Kadlec et al., 2014).

Yet access to advising is vital as nationwide, college students take up to 20 percent more courses than are needed for graduation on average – not motivated by a desire for a diverse curriculum, but because they had to rethink their plans several times. In an environment in which time to degree has considerable implications for a student's likelihood of successfully graduating, a semester of extra coursework plays a crucial factor, especially for students who attend part time, or for whom financial impacts weigh heavily (Complete College America, 2011).

The process of navigating institutions from admission to graduation involves large numbers of crucial decisions, and once again, the information gap plays its part in the achievement gap.

Information and choice clearly have a significant impact on a student's ability to navigate through a degree successfully. But this significantly raises the stakes on the ways in which the information is presented and how the choices are framed. Schwartz (2004) has argued for a paradox of choice – that having too many options can lead to a decision paralysis. Tversky and Kahneman have carefully analyzed how decisions are made in the face of an abundance of choice (Kahneman, 2011; Kahneman & Tversky, 1979; Tversky & Kahneman, 1974). They, and others, have found that when presented with too many choices people fall back on a variety of rules–of–thumb, anecdotal evidence, or rely on cognitive ease and the halo effect. Often, poorer choices are made in situations of an abundance of choice, using these fall back methods, than in situations with more limited choice. In fact the literature on choice overload suggests that too many options can result in several adverse experiences including a depletion of cognitive resources and post–decision feelings of regret (Reed, DiGennaro Reed, Chok, & Brozyna, 2011; Schwartz, 2004). Given the multiplicity of choices entailed in selecting from a college's array of majors or programs, and then satisfying the curricular requirements they require, these adverse experiences may play a significant part in student success, especially for at–risk populations. In fact it seems that a more focused choice structure would be far more effective and preferred (Diamond et al., 2014; Kadlec et al., 2014; Reed et al., 2011; Schwartz, 2004).

While these educational achievement gaps have remained stubbornly present, one promising avenue of attack seems to be the use of predictive analytics to provide individualized information to each student, and so to more evenly level the information playing field.

Predictive analytic techniques move from a retrospective reporting data stance toward the use of large data sets to make detailed predictions about the future. These predictive models enable strategic action to be taken in the present to potentially provide significant improvements in the future. In this vein an appropriately designed system could use the perspective of the past to better inform students, and conversations between students and advisors. Such a system could allow advisors and students to make plans for future semesters, illuminated by the knowledge of courses or even majors in which past students with similar programs, grades and course histories had found success. It could also provide a focused choice architecture in which students could choose from a more limited selection of majors or courses that have been individualized to them, whilst leaving all possibilities available.

Recent Work to Respond to this Challenge

My recent work at Austin Peay State University and now at the Tennessee Board of Regents has, in part, been focused on finding ways to empower student choices by creating choice architectures that improve the information available to each student. The concept was to combine predictive analytics with behavioral economics to create an environment that would help students and advisors select impactful courses. We were intentional in providing an interface that neither restricts nor prescribes their choices, but instead empowers choice by creating an information source with a larger than human viewpoint and supported by data from previous choice patterns (Denley, 2012).

Recommendation systems implemented by companies such as Netflix, Amazon and Pandora are a familiar feature of life today. We decided to create an interface in that vein, and developed a course recommendation system (Degree Compass) that successfully pairs current students with the courses that best fit their talents and program of study for upcoming semesters. The model combines hundreds of thousands of past students' grades with each particular student's transcript to make individualized recommendations for each student. However, the recommendations in this system had to be made within the confines of each student's degree structure, and in a fashion that aligned more closely to the concerns of effective advising if it truly were to level the information field. In contrast to systems that recommend movies or books, these recommendations do not depend on which classes students like more than others. Instead it uses predictive analytics techniques based on grade and enrollment data to rank courses according to factors that measure how well each course might help the student progress through their program. In their 2009 book, Thaler and Sunstein discuss strategies to better structure and inform complex choices (Macfadyen, Dawson, Pardo, & Gasevic, 2014). Degree Compass was designed with this in mind to create a choice architecture to nudge students toward course selections in which the data suggest they would have the most productive success, but using an interface that would minimize choice overload.

While these educational achievement gaps have remained stubbornly present, one promising avenue of attack seems to be the use of predictive analytics to provide individualized information to each student, and so to more evenly level the information playing field.

[2] Degree Compass is now a commercially marketed product, available from D2L Incorporated

The algorithm liaises with the institution's degree audit system to find the courses that would satisfy some as yet unsatisfied degree requirement, if the student were to take that course. From these courses that could apply directly to the student's program of study, the system selects those courses that best fit the sequence of courses in their degree, recommending courses that are curricularly more central before those which are more specialized. That ranking is then overlaid with a model that predicts the courses in which the student will achieve their best grades. In this way, the system most strongly recommends those courses which are necessary for a student to graduate, core to the institution's curriculum and their major, and in which the student is expected to succeed academically.

The recommended course list is conveniently displayed in a web–based interface on the secure side of the institution's information portal. This interactive interface provides information on each recommended course's curriculum and requirements, what role that course plays in the student's degree, as well as class availability in upcoming semesters. The student is able to filter the list to show only classes that are offered online, or face–to–face, or only at particular campuses to refine their decisions according to some practical constraints.

The concept was to combine predictive analytics with behavioral economics to create an environment that would help students and advisors select impactful courses.

The strength to which the system recommends each particular class is communicated by a star rating. A five star class is one that, amongst the presently available courses, best fits the student's curricular constraints, and is one in which the student is predicted to earn as good a grade as they might earn in any other course that would fulfill their requirements. It does not necessarily mean that they will get an A grade. Indeed the interface does not reveal predicted grades to the student. However, all of this information is available to advisors as a tool for academic advising that supplements the information available when providing advice to their advisees.

The interface also provides a majors recommendation system called MyFuture. For a student who has already identified their major, MyFuture provides information about concentration choices and degree pathways, as well as links to prospective career paths, job availability and O*Net statistics for graduates in that major. For a student who is yet to choose a major, or is thinking about changing their major, it provides a list of majors in which that student is predicted to be the most academically successful. Again, for each of these majors, information is provided about concentration choices and degree pathways as well as prospective career paths and job availability. MyFuture uses data–mining techniques to identify the courses that are the best indicators of success in each of the institution's programs – the courses that capture the flavor of each major – and uses Degree Compass' technology to predict course grades and find the majors in which each student will be the most academically successful.

The system was developed in collaboration with faculty, advisor and student input to create an interface that would be able to supplement the advising process. The interface itself was developed to allow commonly utilized functionality in a familiar format. When developing the grade prediction engine for these tools, we chose the data sources on which to base the predictions carefully. Since one of the objectives was to try to impact the performance of subpopulations for which there has been an achievement gap in the past, we chose not to use any demographic information in the model. We also chose to make the system faculty–agnostic by not disaggregating the grading patterns of different faculty. Conversations with faculty members suggested that by doing this there would be greater faculty involvement in the project, and greater utility for the tool.

What the Data Say about the Impact of Degree Compass

We developed a strong assessment structure to assess the impact of Degree Compass on student success (Denley, 2013). Data collected as part of the Degree Compass project fell largely into three categories. First, because courses are recommended to students based on curricular fit, together with a prediction of the grade that student would earn if they were to take the class, it is crucial to collect data that establish the accuracy of the grade predictions. Degree Compass was built to track the predicted grade as well as the earned grade for each student in each semester in each class in which they were enrolled. Secondly, given that advice from Degree Compass is useful only if it is consulted, the system used click–traffic data to provide information about the system's use. Focus groups and surveys also provided feedback about the usability of the interface and other features that users might consider informative. Finally, the aim of the project was to empower students to make more advantageous choices in their education that would help them move effectively through their curriculum. Consequently we measured student success and progression through their curricula.

Instead it uses predictive analytics techniques based on grade and enrollment data to rank courses according to factors that measure how well each course might help the student progress through their program.

Our initial results for the 10,873 students at Austin Peay State University (APSU) were very encouraging. However, it was important to establish that our modeling techniques could calibrate themselves to differing institutional settings and student populations. Generous support from Complete College America and the Bill and Melinda Gates Foundation allowed us to replicate the system at three other schools in Tennessee – two community colleges and one university – adding another almost 40,000 students. Fortunately, the results from all three campuses replicated the ongoing grade prediction resolution achieved at APSU. Data from Fall 2012 showed that the average predicted grades in the university settings were within 0.59 of a letter grade of the awarded grades, and 89 percent of those who were predicted to pass the course indeed passed. In the community college setting, average predicted grades were within 0.64 of the awarded grades, and 90 percent of students who were predicted to pass the course did so. These results confirmed that the grade prediction engine successfully predicts grades in settings across the higher education spectrum, from a rural community college to an urban research university.

Of course, the motivation behind this work was not to predict grades, but rather to provide a choice architecture in which students and advisors could make more nuanced decisions about degree programs. Using Degree Compass as part of academic advising at APSU has steered students towards more classes in which they would more readily succeed, both by passing the course in greater numbers and also achieving higher grades. A comparison of student grades before the introduction of the system with those today shows a steadily increasing ABC%, with grade results across the institution today more than 5 standard deviations better than those in Fall 2010. This very statistically significant shift was apparent across the student body, from freshmen to seniors. We saw similarly significant increases for several subpopulations, including African American students (an increase of 2.1 percent, with 2.89 standard deviations) and Pell recipients (an increase of 3.9 percent, with 7.7 standard deviations). These figures are not results from a sampling of the student population, but include the entire undergraduate student body.

The performance data above clearly demonstrate that students in both the university and community college settings progress more effectively through their degree programs when they follow a course sequence informed by data–analytics.

While it is still early to make general connections between Degree Compass and graduation rates, since the system was introduced at APSU in Spring 2011, the six–year graduation rate has increased from 33 percent to 37.4 percent, with the greater gains for low–income students (increased from 25 percent to 31 percent) and African American students (increased from 28.7 percent to 33.8 percent).

On a more granular level we carried out a detailed analysis of the data to connect Degree Compass recommendations with student successes in their classes and progression through their degrees. Historically, the grade distributions across all four campuses, of all students, showed a picture in which 63 percent of the time a student received an A or a B grade in their course. Using Degree Compass, a much larger proportion of the students who were predicted to earn a B or above were actually awarded that grade. Indeed, on each campus more than 90 percent of students who took a course in which they were predicted to get at least a B actually earned an A or a B grade. The analysis shows that this effect was evidenced at every school and at every course level from developmental classes through upper–division courses.

It is clear that in a model that uses the past to influence the future there is the danger of perpetuating or even reinforcing existing stereotypical trends. However this need not be the case. One of the reasons we chose not to employ demographic information as part of the predictive modeling was precisely to build in safeguards against such phenomena.

For each of the institutions the number of earned credits was highly correlated with number of recommended classes that were part of a student's semester schedule. For instance, those students who took a 12–hour schedule that contained no recommended classes earned only 2.5 credits on average, compared with 10.5 credits for those students whose entire schedule was crafted from recommended courses (see Figure 1). Analysis of other attempted loads showed similar results. With correlation coefficients ranging from 0.7 to 0.9, this connection translates into significant gains when students take recommended classes in comparison with taking classes that are not recommended.

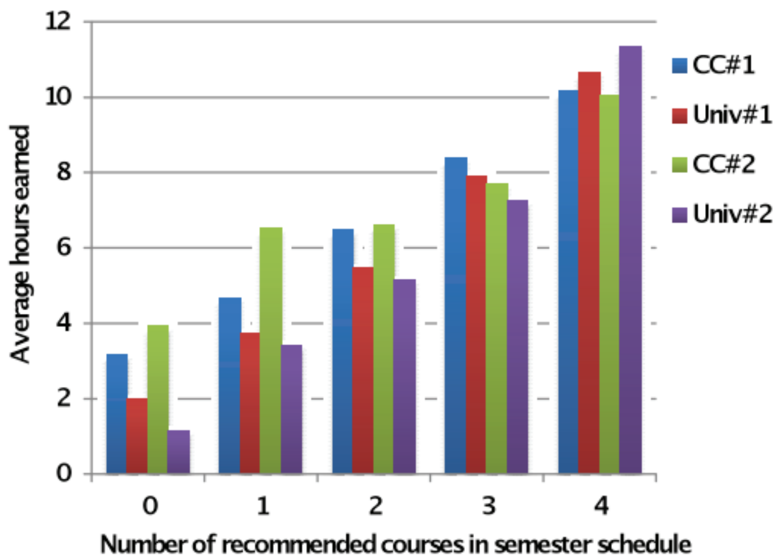


Figure 1. Comparison of average earned hours in a 12–hour schedule disaggregated by the number of recommended classes.

Further analysis of attempted and earned hours revealed that the achievement gap between the average hours earned by white students and average hours earned by African American students reduced significantly for those students who took classes recommended by Degree Compass. For instance among students who attempted 12 hours, white students earned 10.06 hours on average, while their African American peers earned 8.06 hours on average. As we have seen, this is the familiar achievement picture nationally. However, for those students who took 12 hours of courses all of which were recommended by Degree Compass, all students did better, regardless of ethnicity. White students earned 11 hours while African American students earned 10.3 hours on average. The 20 percent achievement gap was more than cut in half (see Figure 2). We see much the same picture for low–income students. Among students who attempted 12 hours, low–income students earned 8.35 hours on average, while their peers earned 10.07 hours on average. However, for those students who took 12 hours of courses, all of which were recommended by Degree Compass, low–income students earned 10.3 hours while their peers earned 11.04 hours on average. Once again, all students did better, and again the achievement gap was cut in half.

Conclusion

Degree Compass has crystalized a number of topics concerning the role that predictive analytics might play in higher education and student success initiatives in particular. First, as a proof of concept, it is now apparent that student success interventions powered by predictive

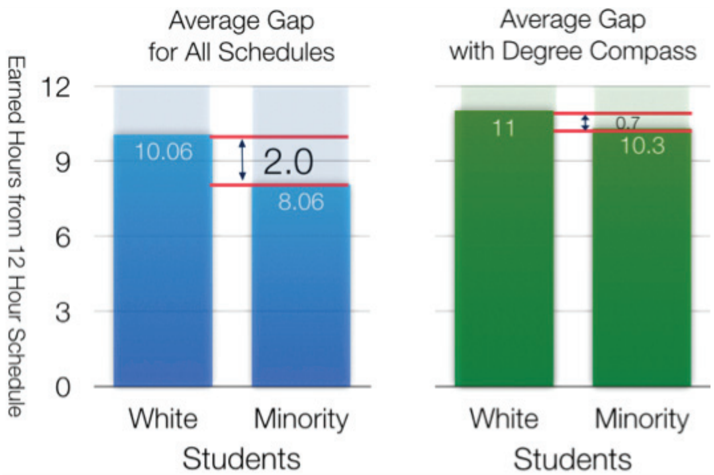


Figure 2. Comparison of average earned hours from a 12–hour schedule disaggregated by race for students in general and students who took only courses recommended by Degree Compass.

analytics are capable of moving the needle on degree completion. The performance data above clearly demonstrate that students in both the university and community college settings progress more effectively through their degree programs when they follow a course sequence informed by data–analytics. Furthermore, there have been precious few approaches that have been able to appreciably close the educational achievement gaps for race and income, and fewer still that can be scaled. Once again, the data suggest that this approach is one that is effective and can be broadly applied at scale.

This approach, however, has highlighted a number of educational issues. It is clear that in a model that uses the past to influence the future there is the danger of perpetuating or even reinforcing existing stereotypical trends. However this need not be the case. One of the reasons we chose not to employ demographic information as part of the predictive modeling was precisely to build in safeguards against such phenomena. The system is designed to be able to use additional data sources as they become available. However, the data that we have collected so far seem to suggest that our current approach has been successful.

In a similar vein, by nudging students towards courses in which they are predicted to have greater success there is the possibility that we may erode academic rigor by systematically steering students towards the easy classes. It may be interesting to contemplate whether when a student takes a class in which they have an increased likelihood of success they are taking an easier class. The experience in the class is as much a function of the student's preparation or talent as it is the challenge of the course. Indeed, as faculty we are all guilty of following the easier route and studying a topic in which we had talent and insight rather than taking the academically more challenging route of choosing a subject for which we had no affinity.

This is not computerized decision making, but technology–informed choice.

One of the important features of Degree Compass is that it only suggests courses that satisfy existing degree requirements. The curriculum is only as rigorous as the courses that can be taken to navigate it, and those remain unchanged. Consequently, the courses that are suggested by the technology are courses that any student might always have chosen and any advisor might always have advised a student to take. The issue comes down to how a student's or advisor's knowledge of the curriculum might inform that choice. It is also an important observation that the suggestions are just that. This is not computerized decision making, but technology–informed choice. The software provides additional information which the student and advisor are then able to use to make more informed decisions. The influence of a plausible default is an important aspect of this, and is an intentional feature of the choice architecture provided in the interface, but the choices that the student and advisor make are still their free choice.

The system only ever suggests courses that satisfy unmet degree requirements. This has the potential to reduce the numbers of excess hours that students currently take. By only suggesting courses that meet degree requirements there is the possibility that the students' experience of the aspect of discovery and intellectual curiosity in the educational process may be stifled. However, transcript analysis shows that more often than students choosing courses off their curricular path because of intellectual curiosity, they actually take these classes simply because the course they would like to choose is unavailable. Since the data now clearly support that students taking the courses that they need is a crucial aspect of student success, it is incumbent on us to offer the classes that students need, when they need them. If we employ predictive technology to ensure that the skeletal structure of the degree is seamlessly available to students, we create the flexibility for more intellectual curiosity should the student choose.

In fact, a deep dive into data at the Tennessee Board of Regents has allowed me to create strategic insights into the structure of the system and how students succeed and fail. These insights are being used to inform changes to system policy, as well as direct broad–scale system initiatives.

Here we have concentrated on seeing how individualized analytics can be used to help optimize course and curricular selections, but there are many other ways in which these kinds of technology can be utilized across higher education. This work demonstrates how predictive analytics can provide a larger–than–human viewpoint that can inform student choice. We are starting to see how these kinds of recommending systems can empower decisions by program coordinators, and institutional leadership. In fact a deep dive into data at the Tennessee Board of Regents has allowed me to create strategic insights into the structure of the system and how students succeed and fail. These insights are being used to inform changes to system policy, as well as direct broad–scale system initiatives. It seems likely that over the coming years we will see more and more ways in which predictive analytics and data–mining technology coupled with behavioral economics will play roles in higher education on every scale (Johnson et al., 2013; cf. O'Reilly & Veeramachaneni, 2014).

References

Chen, X., & Carroll, C. D. (2005). *First generation students in postsecondary education: A look at their college transcripts.* NCES 2005–171. Washington, DC: National Center for Educational Statistics.

Complete College America. (2011). *Time is the enemy*. Washington, DC: Author.

Denley, T. (2012). Austin Peay State University: Degree Compass. In D. G. Oblinger (Ed.), *Game changers: Education and information technologies* (pp. 263–267). Louisville, CO: EDUCAUSE.

Denley, T. (2013). Degree Compass: A Course Recommendation System. *EDUCAUSE Review Online*. Retrieved from http://www.educause.edu/ero/article/degree–compass–course–recommendation–system

Diamond, A., Roberts, J., Vorley, T., Birkin, G., Evans, J., Sheen, J., & Nathwani, T. (2014). *UK review of the provision of information about higher education: Advisory study and literature review*. Leicester, UK: HEFCE.

Hagelskamp, C., Schleifer, D., & DiStasi, C. (2013). *Is college worth it for me?: How adults think about going (back) to school*. New York, NY: Public Agenda.

Jenkins, D., & Cho, S–. W. (2012). Get with the program: Accelerating community college students' entry into and completion of programs of study. New York, NY: Community College Research Center.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H. (2013). NMC horizon report: 2013 higher education edition. Austin, TX: The New Media Consortium.

Kadlec, A., Immerwahr, J., & Gupta, J. (2014). *Guided pathways to student success: Perspectives from Indiana college students & advisors*. New York, NY: Public Agenda.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*(4), 263–292.

Kelly, P. J. (2005). *As America becomes more diverse: The impact of state higher education inequality.* Boulder, CO: National Center for Higher Education Management Systems.

Kirst, M. W., & Venezia, A. (Eds.). (2004). *From high school to college: Improving opportunities for success in postsecondary education*. San Francisco, CA: Jossey–Bass.

Lumina Foundation. (2014). *A stronger nation through higher education*. Indianapolis, IN: Author.

Macfadyen, L. P., Dawson, S., Pardo, A., & Gasevic, 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.

National Association of System Heads (NASH), & The Educational Trust. (2009). *Charting a necessary path: The baseline report of public higher education systems in the Access to Success Initiative*. Washington, DC: Authors.

National Center for Education Statistics (NCES). (2012). *Higher education: Gaps in access and persistence study.* Washington, DC: Author.

O'Reilly, U. M., & Veeramachaneni, K. (2014). Technology for mining the big data of MOOCs. *Research & Practice in Assessment, 9*(2), 29-37.

Reed, D. D., DiGennaro Reed, F. D., Chok, J., & Brozyna, G. A. (2011). The "tyranny of choice": Choice overload as a possible instance of effort discounting. *The Psychological Record, 61*(4), 547–560.

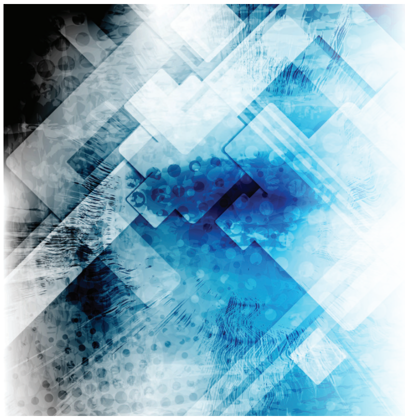
Schwartz, B. (2004). *The paradox of choice: Why more is less*. New York, NY: HarperCollins.

Smith, J. S., & Wertlieb, E. C. (2005). Do first–year college students' expectations align with their first–year experiences? *NASPA Journal, 42*(2), 153–174.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. New York, NY: Penguin Group.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, New Series, 185*(4157), 1124–1131.

U. S. Census Bureau. (2006). *Current population survey, October supplement, 1972–2006*. Washington, DC: Author.



Abstract

Civitas Learning was conceived as a community of practice, bringing together forward-thinking leaders from diverse higher education institutions to leverage insight and action analytics in their ongoing efforts to help students learn well and finish strong. We define insight and action analytics as drawing, federating, and analyzing data from different sources (e.g., ERP, LMS, CRM) at a given institution to produce deep predictive flow models of student progression and completion coupled with applications (apps) that take these data and bring them to advisors, students, faculty, and administrators in highly consumable/ useable ways. Through three case studies, this article provides a closer look at this iterative work unfolding in diverse institutions, addressing diverse student success challenges, and achieving significant positive results on student progression. The article underscores a key finding: there is not a one-size-fits-all predictive model for higher education institutions. We conclude with a discussion of key findings from these cases and observations to inform future related work.

AUTHORS

Mark David Milliron, Ph.D.
Civitas Learning

Laura Malcolm, M.A.
Civitas Learning

David Kil, M.S.E.E., M.B.A.
Civitas Learning

Insight and Action Analytics: Three Case Studies to Consider

Civitas Learning was conceived as a community of practice, bringing together forward–thinking leaders from diverse higher education institutions to leverage insight and action analytics in their ongoing efforts to help students learn well and finish strong (Fain, 2014; Thornburgh & Milliron, 2013). Our fast–growing community of practice now includes more than 40 institutions and systems, representing more than 570 campuses, serving more than 1.45 million active students. It includes research one institutions, emerging research and access universities, independent colleges, community colleges, and private sector universities. We work with cross–functional groups of administrators, IT teams, IR teams, advisors, and faculty members, most of whom are leading large–scale student learning and completion programs, often catalyzed by federal, state, foundation, and institutional dollars, pressures, and aspirations. Some initiatives include the Obama Administration 2020 Goals (Higher Education, 2014), Complete College America (2014), Bill & Melinda Gates Foundation Postsecondary Initiative (Postsecondary success strategy overview, 2014), Lumina Foundation for Education's Goal 2025 (Lumina Foundation Strategic Plan, 2013), Texas Student Success Council (2014), Hewlett Foundation's Deeper Learning Initiative (2014), and Kresge Foundation's Education Initiative (2014; Milliron & Rhodes, 2014). It is important to note that we do not conceive of our work as another new initiative. Indeed, many of these institutions report that they are already reeling from “initiative fatigue.” Rather, our insight and action analytics infrastructure is meant to be a powerful resource to try, test, and power deeper learning and student success initiatives (Kim, 2014).

CORRESPONDENCE

Email
mark.milliron@civitaslearning.com

We define insight analytics as the family of activities that bring data from disparate sources together to help create a more complete view of student progression. In the most basic terms, this means (a) federating data from an institution’s Student Information System (SIS) and Learning Management System (LMS); (b) using sophisticated data science tools and techniques, including machine learning, data availability segmentation and clustering, to create and compete feature variables derived from the diverse sources; (c) building an array of predictive models; and then (d) leveraging a variety of visualization techniques we explore the resulting historic and predictive student progression/flow models for insights that help better understand how students succeed and face challenges on their higher education journeys. Once the models are developed, we create a cloud–based, production–quality, predictive–flow–model infrastructure for each institution that is updated at minimum on a rolling five–term cadence to keep the student–level predictions as current as possible. From here, more sophisticated insight analytics work includes adding additional data sources in this mix, such as Census, application data, card swipe, CRM, and more, and then testing these new data streams for added predictive power to drive decisions about how or whether to add them to the production system. See the Appendix for a deep dive on some of these techniques.

We created a platform application called Illume™ that brings insights from this work to our institutional partners, allowing them to view student progression dynamics filtered by chosen segments (e.g., part–time, full–time, Pell recipients, distinct campuses, members of intervention category), often testing assumptions about performance and possible historic and predictive trends (Figure 1.1). The application also surfaces powerful predictors for distinct segments, which are feature and point variables contributing significantly to the success or challenge of a given segment. For example, a feature variable we derive called affordability gap – the delta between financial aid received and tuition owed – is often a far more powerful predictor for first–time students than placement test scores. The diverse segment and cluster analyses often point to relationships that are non–intuitive or surprising, and other times reaffirm long–held assumptions. Either way, they are useful in starting conversations about tipping points, momentum points, and possible dynamics at work in systems, processes, policy, and practice at the institution.

It is important to note that we do not conceive of our work as another new initiative. Indeed, many of these institutions report that they are already reeling from "initiative fatigue."

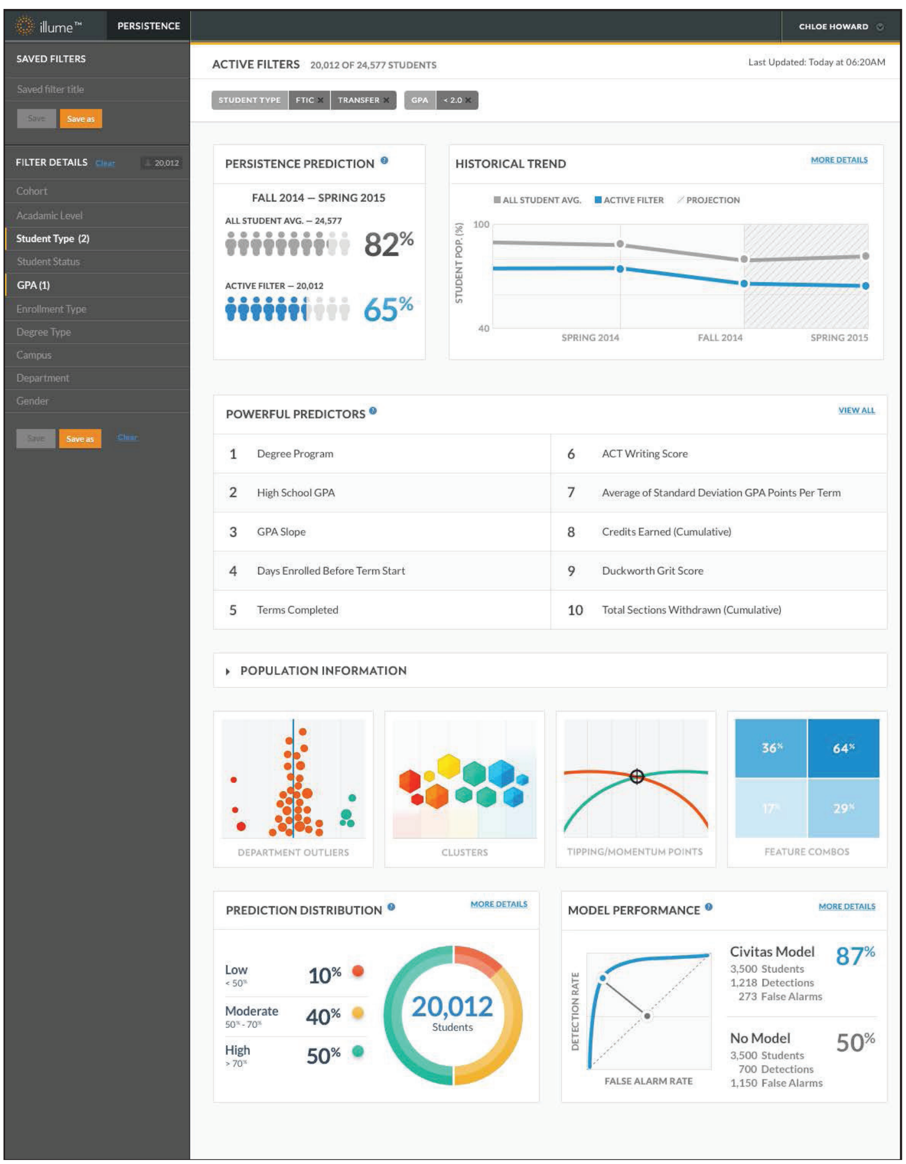


Figure 1.1

This insight analytics infrastructure can be useful, to be sure. But in our work over the last three years we have found that this predictive flow platform is more a predicate than a solution. The insights derived can make a stronger impact on student success when used to power action analytics. Action analytics include applications (apps) that use design thinking, user–interface disciplines, and strategic workflow to leverage insight analytics in easy to consume, engaging, and highly useable formats to help administrators, advisors, faculty, and even students interact with these data to help understand risk and success factors, target and test interventions, and guide choices, outreach, and activity. We have developed a family of action–analytic apps that include our Degree Map™, Inspire™, and Hoot.Me™ family of apps (Figure 1.2). Each of these is being deployed at different institutions and are being tried, tested, and tuned as the work of learning about how to bring insight and action analytics into the daily operations of institutions continues.

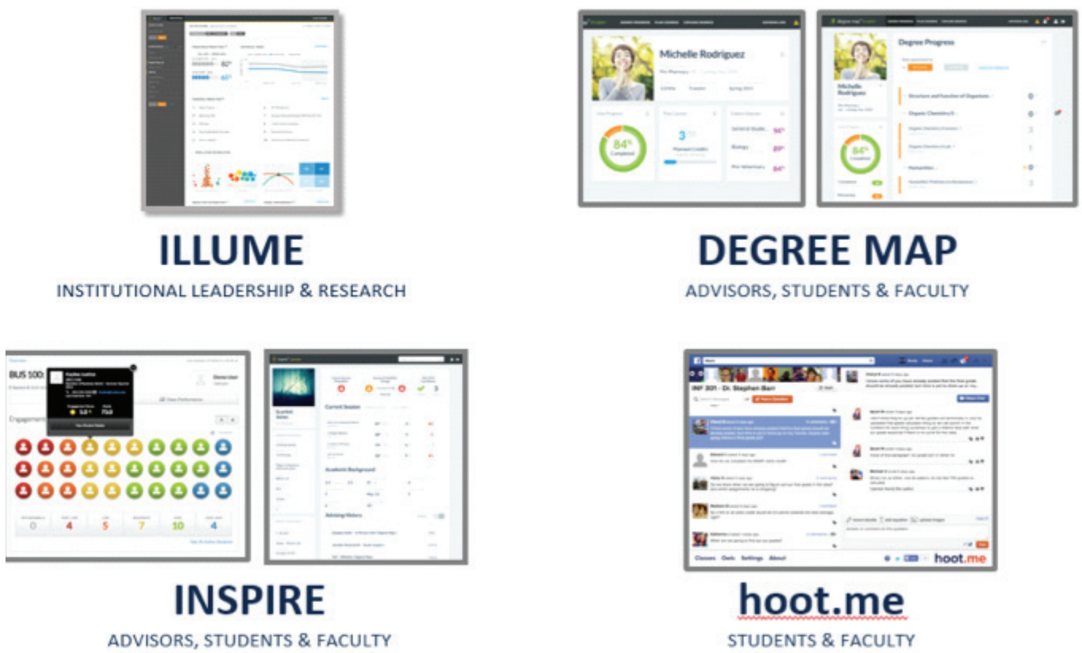


Figure 1.2

Rather, our insight and action analytics infrastructure is meant to be a powerful resource to try, test, and power deeper learning and student success initiatives.

There is, of course, an array of learning–centered and student–completion–centered action applications at work in the field of higher education, from basic early–alert systems to comprehensive CRM tools (Blumenstyk, 2014; Milliron, 2013). However, most of these have choice architectures and engagement tools powered by heuristic triggers and set configurations as opposed to institution–specific, student–level predictive flow models. Others leverage quite sophisticated advanced analytics, but only in the context of their application (e.g., several adaptive learning tools). However, many of these action applications are likely to add insight–analytic linkages on the road ahead and will move into a growing ecosystem of what we call Action Analytic Applications. Indeed, we are likely to see dozens, if not hundreds of these, emerge in the months and years ahead.

It is important to note that these action analytic applications can be data streams in and of themselves that can inform and improve the insight analytics work, creating an ongoing and continuously improving analytics infrastructure. For example, both the Inspire for Advisors and Inspire for Faculty Apps generate data on tried interventions with different students that can inform future suggestions for advisors and faculty members. Hoot.me, which is a crowd–sourced, student–driven, question–and–answer community app generates engagement and social interaction data. Indeed, some future action analytic application may be used primarily to generate data – e.g., an app that gathers wellness behaviors or non–cognitive mindsets through micro surveys.

The interplay between and the process of learning more about insight and action analytics has been at the heart of our work for the last three years. The community of practice site, Civitas Learning Space, showcases the ongoing initiatives in an effort to inform and engage a broader audience. Moreover, the Civitas Learning partner community comes together twice a year for summits on data science, intervention strategies, and future planning (Rees, 2014).

What follows is a closer look at three of our partner institutions as they brought together their insight and action analytics initiatives. We present three cases in an effort to show how this iterative work unfolds in diverse institutions, approaching diverse student success challenges, and to underscore a key finding: There is not a one–size–fits–all predictive model for higher education institutions. Each institution has its own predictive student flow and leaders, teachers, and advisors need to understand and engage their student success strategies in the context of their own students, policies and practices. We will come back in the concluding section to offer observations for those interested in learning more or joining in similar efforts.

Case Study One: Early Intervention for Course Success

Executive Summary

Leveraging Civitas Learning's Illume predictive analytics platform and Inspire application for administrators and advisors, Partner Institution A ran a pilot program to test the efficacy of using predictive–analytics–based interventions on driving improvements to student course completion rates. Over the course of three terms starting in Spring 2013, predictive models were built, approaches to intervention were tested, and outcomes were evaluated using a randomized test and control pilot approach. In the first two terms of the pilot, no statistically significant improvements to outcomes were measured. In Fall of 2013 with a pilot group of ~14,000 enrollments (~7,000 each in test and control) and applying learnings from previous terms, the institution realized an average improvement of 3% at a 98% confidence level for statistical significance test vs. control. This translates into 210 student enrollments that successfully completed their course that otherwise would have failed or withdrawn.

We define insight analytics as the family of activities that bring data from disparate sources together to help create a more complete view of student progression

Introduction

Institution A is a 4–year access institution with greater than 50,000 students including undergraduate and graduate. They offer on–campus programs and courses as well as online programs through an online campus.

The focus of the pilot with Institution A was using advanced analytics to understand student risk, the variables that contribute most to student success, and most importantly how to make these insights actionable to improve student outcomes. Ultimately, the institution goal is a more personalized student experience and a better probability for student success, which translates to higher course completion, retention, and graduation rates to fulfill their institutional mission.

Methodology

Three pilots were conducted over the course of three terms (Spring 2013, Summer 2013, and Fall 2013) using randomized assignment of all enrollments within a section to test or control groups. While random assignment at the enrollment level would be preferred to reduce selection bias based on section and instructor, operational constraints prevented this approach.

In order to evaluate the potential section level bias, baseline predictions of course success were used to evaluate whether the sections were biased. Deltas between prediction of course success showed no statistically significant difference between test and control group in terms of enrollment likelihood to successfully complete.

In all three pilots course success was defined as finishing the course with a grade of C or better for undergraduate enrollments and B or better for graduate enrollments. For outcomes analysis, statistical significance was computed using Fisher's exact test, widely used in the analysis of contingency tables (Fisher, 1954).

In Spring 2013, nine courses (four graduate and five undergraduate) participated in the pilot with 2,279 enrollments in total. In Summer 2013, the pilot grew to ten courses (five graduate and five undergraduate) and 6,832 enrollments. Finally, in Fall 2013 the pilot included 15,500 enrollments across 25 courses (10 graduate and 15 undergraduate).

Study

While predictive analytics have the potential for wide applicability across the student lifecycle, the starting point for this pilot focused where there could be concrete results that could be measured in a short amount of time – student successful course completion.

Pilot goals were:

- Demonstrate that predictive analytics in combination with targeted interventions can improve student outcomes.
- Evaluate which interventions produce better outcomes.
- Learn from the process and determine strategies to scale predictive analytics for personalized interventions.

For example, a feature variable we derive called affordability gap – the delta between financial aid received and tuition owed – is often a far more powerful predictor for first–time students than placement test scores. The diverse segment and cluster analyses… are useful in starting conversations about tipping points, momentum points, and possible dynamics at work in systems, processes, policy, and practice at the institution.

Pilot roll–out. Leveraging historical data, Civitas Learning developed institution specific predictive models to evaluate the complex set of variables contributing to student success. These models provide an individualized risk prediction of each student's likelihood to successfully complete a course, with greater than 80% accuracy prior to the course start. As student behaviors were introduced into the models over the course term, the student's risk prediction was continually updated, providing an increasingly accurate measure of course completion likelihood.

Civitas Learning's Inspire application delivered these predictions in an actionable way to academic administrators and advisors, so that they could understand which enrollments were at–risk and apply timely interventions and support. Users analyzed data, segmented student populations and implemented targeted communications directly from the application.

Spring 2013 pilot. In the initial Inspire for Administrators roll out in the Spring of 2013, based on insights from the application, subgroups were analyzed to determine variance in probability to succeed based on many predictive factors (including GPA, attendance patterns, grades, terms of enrollment, course credits and many more). Email communications were sent from the Inspire application by academic program administrators based on student risk factors. Content of the emails was determined by the program administrator and varied across programs. Fifty–one percent of enrollments received an email intervention with an average of 1.71 interventions per enrollment. The control group did not receive interventions.

Summer 2013 pilot. In the Summer of 2013, using the same predictive model, academic program administrators expanded the pilot to a larger number of courses and enrollments. Again, email communications were sent from the Inspire application by program administrators based on student risk factors. However, in this pilot, the test group was broken into four sub–groups to test varied outreach approaches including templatized content and timing differences. Outreach approaches for each test group were developed by a committee of academic leads from across programs. In the Summer pilot, 54% of enrollments received an email intervention with an average of 1.36 interventions per enrollment. The control group did not receive interventions.

Fall 2013 pilot. Deployment and experimentation with selected interventions allowed for early testing of intervention approaches during the spring and summer terms. Processes were operationalized and refined, and best practices were established regarding the dissemination of interventions in preparation for the Fall 2013 term.

In Fall of 2013, Academic Program Administrators and Advisors used the app (Inspire for Administrators) to determine students most in need of intervention, then pulled from a prepared suite of intervention tools, messaging, emails, and calendar items to provide support in a timely, empathetic, appropriate way to students in the test group. The control group did not receive interventions.

Findings

Model performance. Looking retroactively at model performance, at an individual student level, predictive models were able to identify with 83% accuracy on the first day of a course the students who would successfully complete and by day seven the accuracy level moved to 86%. Model performance remained at these levels across the three pilots.

Outcome performance. In Spring of 2013, the test group outperformed the control group in successful course completion by 122 basis points. However, the p–value was 0.2677 not reaching statistical significance. Institution A found these results to be promising and developed a series of templatized outreach plans to facilitate outreach for the next term.

In Summer 2013, there was no measurable impact on successful course completion. Theories as to why there was no improvement focused on the complexity of the intervention outreach plans and the user base of the application. Institution A decided to simplify the outreach approach for fall and to add advisors to the pilot to assist with student outreach.

In Fall of 2013, the test group of ~5,000 undergraduate students outperformed the control group in successful course completion by 300 basis points. This result had a p–value of 0.05 reaching statistical significance at a confidence level of 95%. There was no measurable improvement for graduate students.

We present three cases in an effort to show how this iterative work unfolds in diverse institutions, approaching diverse student success challenges, and to underscore a key finding: There is not a one–size–fits–all predictive model for higher education institutions. Each institution has its own predictive student flow and leaders, teachers, and advisors need to understand and engage their student success strategies in the context of their own students, policies and practices.

Case Study Two: Early Intervention by Faculty for Persistence Gains Executive Summary

Leveraging Civitas Learning's Illume™ predictive analytics platform and Inspire for Faculty application, Partner Institution B ran a pilot program to test the efficacy of using predictive analytics based interventions to drive improvements in student persistence rates. Over the course of three terms starting in Fall of 2012, predictive models were built, an application was launched to facilitate faculty outreach, and outcomes were evaluated. A pilot was conducted across two terms beginning in the Winter 2013 term. During the pilot, faculty used a "heat map" of student engagement to identify and prioritize students for intervention outreach. In the first term of the pilot no statistically significant improvement to outcomes was measured. In the Spring Term of 2014 with a group of ~68,000 online enrollments and applying learnings from previous terms, the institution realized statistically significant persistence improvements.

Introduction

Institution B is a 4–year access institution with more than 20,000 students including both undergraduate and graduate programs. They offer on–ground programs as well as an online campus. The focus of the pilot with Institution B was to use advanced analytics to understand online student risk of successful course completion and persistence and use that understanding for the prioritization and differentiation of outreach by faculty.

Methodology

Two pilots were conducted over the course of two terms (Winter 2013 and Spring 2014). The first pilot focused on undergraduate online students in six high enrollment courses. In the first term, randomized assignment of students to test and control groups created the pilot group. In the second term, because of operational challenges in administering interventions to only test students, propensity score matching was used to identify a matching control group. This allowed for all online enrollments to be in the test group while identifying the control group from historical enrollments.

Propensity-score matching (PSM) is used in observational studies when there is no randomized control group. Simply put, PSM compresses salient features (x) of pilot participants into a single variable called propensity score. It then computes the propensity scores of non-participants using their attributes and finds matching cohorts, such that $p(z=1/x) = p(z=0/x)$, where z is the binary participation variable. This assures that the matching cohorts are statistically similar to the pilot group in x. As an extra security layer, top features (x) from the predictive models (y = f(x)) are used in PSM. This ensures that the created control group is virtually indistinguishable from the pilot group from an outcomes (y) perspective. That is, $p(y/x, z=1) = p(y/x, z = 0)$.

In all three pilots, persistence was defined as re–enrolling in the next term and staying enrolled past the add–drop/census period in the following term. For outcomes analysis, statistical significance was computed using Fisher's exact test, widely used in the analysis of contingency tables (Fisher, 1954).

In Fall of 2013…the institution realized an average improvement of 3% at a 98% confidence level for statistical significance test vs. control. This translates into 210 student enrollments that successfully completed their course that otherwise would have failed or withdrawn.

Courses participating in the pilot grew to all online courses in the second term. The student enrollment count in the Winter term was approximately 15,000 (with 7,500 each in test and control). However, in the Spring 2013 term due to including all online enrollments the pilot grew to ~68,000 enrollments each in test and control groups.

Study

While predictive analytics has many applications, this pilot focused on leveraging faculty outreach to drive improvements to student persistence through effective outreach.

Pilot goals were:

- Demonstrate that predictive analytics, in combination with targeted interventions, can improve student outcomes.
- Focus faculty on improving student engagement in online courses
- Learn from the process and determine strategies to scale predictive analytics for personalized interventions.

Pilot roll–out. Leveraging historical data, Civitas Learning developed institution–specific predictive models to evaluate the complex set of variables contributing to student successful course completion and engagement in online courses. These models provided an individualized risk prediction of each student's likelihood to successfully complete the course. From this model the online course behaviors predictive of course success were identified and used to create a student engagement score. The engagement score was based on a zero to ten point scale and was relative – comparing engagement to all other enrollments taking the same course at the same time. The engagement score weighted behaviors based on their contribution to the predictive model.

Looking retroactively at model performance, at an individual student level, predictive models were able to identify with 83% accuracy on the first day of a course the students who would successfully complete and by day seven the accuracy level moved to 86%.

Civitas Learning's Inspire application then delivered the engagement score in an actionable way to online faculty, so that they could prioritize and differentiate intervention outreach to students. In addition to the engagement score, key information was included to help faculty understand why students were at risk so they could apply timely interventions and support. Using the application, faculty emailed students to drive increased online course engagement. All outreach was tracked so approaches and timing could be analyzed for effectiveness. In addition, since engagement scores were relative, faculty could monitor their section engagement in order to see how their students were doing on engagement compared to the whole.

Winter 2013 pilot. In the initial pilot, conducted during the Winter 2013 term, the predictive models generated a daily engagement score for each student in each section. Faculty used this score to assist in prioritizing outreach for students in the test group. The interface provided direct access to their assigned sections and students as well as the ability to segment students for outreach based on parameters such as current grade in course, engagement score, etc. In addition, the interface allowed faculty to track interventions and see a record of all emails sent to a student.

Finally, a tracking dashboard was deployed that allowed faculty to track week to week progress on engagement, successful course completion and continuation and compare that progress between their section and all other sections of the same course. Faculty used this prediction to assist in prioritizing re–enrollment and differentiating outreach for students in the test group. Faculty used their standard instructional process for control group sections.

Spring 2013 pilot. In the Spring 2013 term, the application was enhanced to allow faculty to "bulk" email students. Bulk email provided faculty the means to email the same content to multiple students, with name personalization, in one action. In addition, "Recommended Outreach" was added to the interface to provide quick links to faculty to assist completion of the most common interventions. For example, one recommendation filtered "students with low engagement who haven't had outreach in the past week" and let faculty email them in one click.

Findings

Model performance. Looking retroactively at model performance by reviewing engagement scores in comparison to final grades, the data show that the scores were highly reflective of successful course completion.

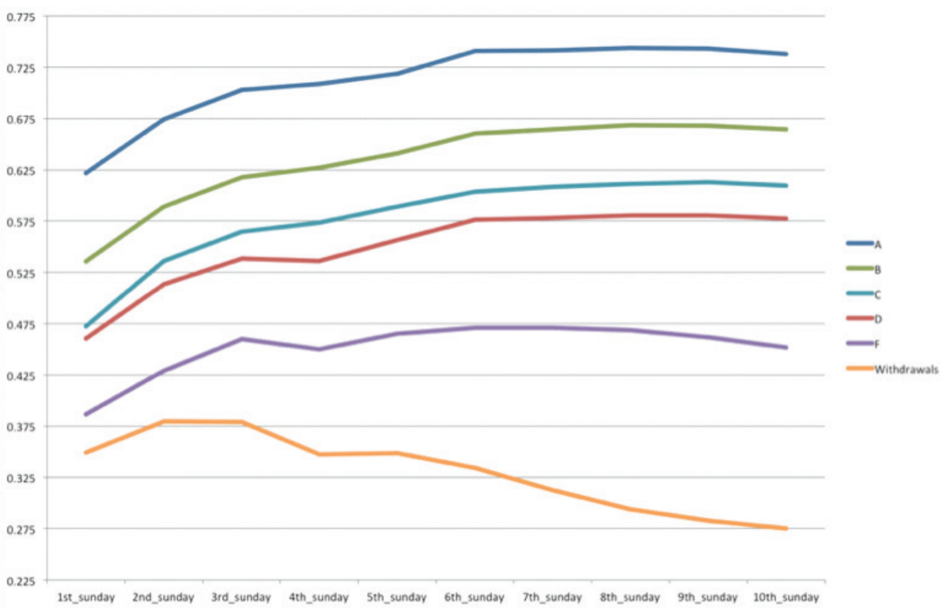


Figure 1.3. Illustration of the week by week engagement score trend in comparison with the student final grade shows the engagement score is highly correlated with successful course completion.

Outcome performance. In the Winter 2013 term, the test group outperformed the control group in persistence by 91 basis points. The result was not statistically significant reaching a p–value of 0.19 with a confidence level of 81%. However, institution B found these results to be promising and in the following term made plans to widen the pilot to include all online courses.

In Spring 2013, persistence rates from the Spring Term into the Summer Term were 321 basis points greater for test group than the control group. This result had a p–value of 0.05 reaching statistical significance at a confidence level of 95%. This result was calculated using retrospective propensity score matching to identify the control group. In order to validate the results a second analysis was done using time–series forecasting and the results held at a statistically significant level.

In reviewing the intervention data by terms completed, for early term students, phone calls where the advisor spoke to the student were the most effective intervention. Conversely, for students with greater than ten terms completed at the institution, email appears to be the best initial intervention.

Case Study Three: Early Intervention by Advisors for Persistence Gains Executive Summary

Leveraging Civitas Learning's Illume predictive analytics platform and Inspire application for Advisors, Partner Institution C ran a pilot program to test the efficacy of using predictive analytics based interventions on driving improvements to student persistence. Over the course of three terms, starting in January of 2014, predictive models were built, approaches to advisor led intervention were tested, and outcomes were evaluated using a randomized test and control pilot approach. In the first two terms of the pilot no statistically significant improvements to outcomes were measured. However, in the May 2014 term with a pilot group of ~10,000 students, and applying learnings from previous terms, the institution realized statistically significant improvements in persistence for students in their first nine terms. Largest gains were realized for new students, with a 762 basis point improvement in persistence when comparing the test to the control group.

Introduction

Insight analytics that are developed using institution–specific data sources – particularly student–level SIS and LMS data – are vital to understanding student flow, as well as targeting and personalizing intervention and outreach.

Institution C is a career–focused 4–year access institution with more than 40,000 students including both undergraduate and graduate programs. They offer on–ground campus locations as well as an online campus. The focus of the pilot with Institution C was to use advanced analytics to understand student risk of re–enrollment and persistence. In addition, the pilot was designed to use that understanding for the prioritization and differentiation of enrollment services through their advising function.

Methodology

Three pilots were conducted over the course of three terms (January 2014, March 2014, and May 2014). The pilot focused on undergraduate online students in six degree programs. In the first two terms, randomized assignment of students to test and control groups created the pilot cohort. In the third term, because of operational challenges in administering interventions to only test students, propensity score matching was used to identify a matching control group. This allowed for all students within the specified degree programs to be in the test group while identifying the control group from other degree programs.

Propensity-score matching (PSM) is used in observational studies when there is no randomized control group. PSM compresses salient features (x) of pilot participants into a single variable called propensity score. It then computes the propensity scores of non-participants using their attributes and finds matching cohorts, such that $p(z=1/x) = p(z=0/x)$, where z is the binary participation variable. This assures that the matching cohorts are statistically similar to the pilot group in x. As an extra security layer, top features (x) from the predictive models ($y = f(x)$) are used in PSM. This ensures that the created control group is virtually indistinguishable from the pilot group from an outcomes (y) perspective. That is, $p(y/x, z=1) = p(y/x, z = 0)$.

In short, there is not a global predictive model that works across institutions with any level of accuracy. You need to “turn the lights on” in your institution.

In all three pilots, persistence was defined as re–enrolling in the next term and staying enrolled past the add–drop/census period in the following term. For outcomes analysis, statistical significance was computed using Fisher’s exact test, widely used in the analysis of contingency tables (Fisher, 1954). Degree programs participating remained consistent across the three pilots. The student count in the January and March terms was approximately 5,000 (with 2,500 each in test and control). However, in the May 2014 term, due to including all students in the selected degree programs, the pilot grew to ~10,000 students with 5,000 each in the test and control groups.

Study

While predictive analytics has many applications, this pilot focused on using predictive analytics to maximize the effectiveness of advising resources in driving re–enrollment and student persistence.

Pilot goals were:

- Demonstrate that predictive analytics, in combination with targeted interventions, can improve student outcomes.
- Maximize application of advising resources to improve persistence.
- Evaluate which intervention approaches produce better outcomes and for which students.
- Learn from the process and determine strategies to scale predictive analytics for personalized interventions.

Pilot roll–out. Leveraging historical data, Civitas Learning developed institution–specific predictive models to evaluate the complex set of variables contributing to student persistence. These models provide an individualized risk prediction of each student’s likelihood to persist at the institution. As student behaviors were introduced into the models over the course term, the student's risk prediction continually updated, providing an increasingly accurate measure of persistence likelihood for advisors.

Civitas Learning's Inspire application delivered these predictions in an actionable way to advisors (student success coaches), so that they could prioritize and differentiate re–enrollment outreach to students. In addition to the prediction, key information was included to help advisors understand why students were at risk so they could apply timely interventions and support. Using the application, advisor managers analyzed data, designed outreach approaches, and assigned advisors to students for intervention. All outreach was tracked so it could be analyzed for effectiveness.

January 2014 pilot. In the initial pilot, conducted during the January 2014 term, institution–specific predictive models were used to generate a "Day 0" report that identified students' probability to persist into the following term starting the day before the new term. This model used student information system (SIS) data to make the prediction. Advisors used this prediction to assist in prioritizing re–enrollment and differentiating outreach for students in the test group. Advisors used their standard re–enrollment process for control group students.

Bringing insight analytics together with action analytics is essential to "moving the needle" on student success. Better precision of the models helps target outreach and improve impact of instruction and advising support.

A probability score between 0 and 1 was generated for each student and students were distributed into five persistence groups (quintiles) based on this score. Groups ranged from very high to very low probability to persist. Advisors were provided with the group assignment for each student along with key academic background information for context. Background information differed depending on whether students were new or continuing.

The report was delivered in the form of a spreadsheet to advisor managers who used it to make advisor assignments and design outreach approaches. Advisors used a combination of email and phone call outreach to test group students. Across the term, re–enrollment was tracked and reported to the group on a weekly basis.

March 2014 pilot. In the March 2014 term, the predictive models were enhanced to include learning management system (LMS) data. In addition, delivery of the spreadsheet moved from a one–time report to a report delivered nightly. As in the January pilot, advisors used this prediction to assist in prioritizing re–enrollment and differentiating outreach for students in the test group. Again, advisors used their standard re–enrollment process for control group students.

May 2014 pilot. In the May 2014 term, the report was replaced by the Inspire for Advisors application which provided a user interface for each advisor to manage their student caseload. The interface provided direct access to their assigned student list as well as the ability to segment students for outreach based on parameters such as degree program, new vs. continuing status, probability group, and recent changes to their probability. In addition, the interface allowed advisors to track interventions and see a record of all outreach administered to a student. Finally, a re–enrollment tracking dashboard was deployed that allowed advisor managers to track week to week progress on continuation and compare that progress between the test and control groups. As in the previous pilots, advisors used this prediction to assist in prioritizing re–enrollment and differentiating outreach for students in the test group. Again, advisors used their standard re–enrollment process for control group students.

Findings

Model performance. Looking retroactively at model performance by reviewing the probability group assignments, the data show that the predictions were highly reflective of actual student persistence rates. For example, for students in the 0–40% probability of persistence range, average actual persistence was 27%. On the other end of the spectrum, for students in the 80–100% probability of persistence range, average actual persistence was 86%.

Figure 1.4 shows the actual Receiver Operating Characteristic (ROC) curves for Institution C to explain salient concepts. Assuming the intervention outreach capacity of 10K students, using the purple model (test) provides 141% improvement (20.9% to 50.5%) in correctly identifying eventual non–persisting students for intervention in comparison to randomly reaching out to students.

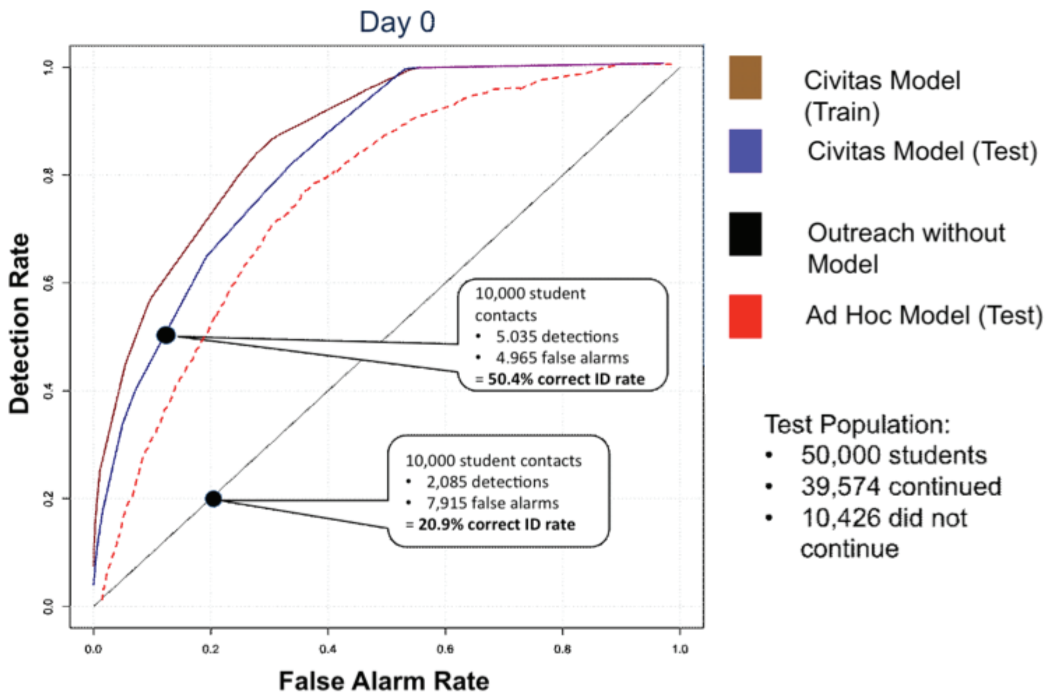


Figure 1.4. The day–0 ROC curves for the final train/test models using data–availability segmentation and clustering, an ad hoc model, and the random chance line.

Outcome Performance

How you bring data to the front lines of learning – e.g., to advisors and faculty – has a significant impact on the effectiveness of these efforts. Modality, timing, visualization, and operational tools matter.

January 2014. The test group outperformed the control group in persistence by 120 basis points. However, the *p*–value was 0.22, not reaching statistical significance. Institution C found these results to be promising and in the following term made plans to operationalize a daily prediction report.

March 2014. There was no measurable impact on persistence in the March 2014 term. Theories as to why there was no improvement focused on the operational complexity of managing a nightly report and distributing assignments to advisors in a timely fashion. Development of an application interface for advisors was underway and became the highest priority for the next pilot.

May 2014. Among new students, persistence rates from the May term into the July term were 762 basis points greater for test group than the control group. This result had a *p*–value of 0.02, reaching statistical significance at a confidence level of 98%. There was no measurable improvement for students past the ninth term of enrollment. Positive, statistically significant improvement was seen for students in their second until seventh term, into their eighth term.

In addition, intervention approaches were analyzed by student persistence probability and also by terms completed. For "Low" and "Moderate Persistence Probability" students, phone calls where the advisor "spoke to" the student were the most effective intervention approach. However, for "High Persistence Probability" students, "spoke to" was only slightly better than an email intervention.

In reviewing the intervention data by terms completed, for early term students, phone calls where the advisor spoke to the student were the most effective intervention. Conversely, for students with greater than ten terms completed at the institution, email appears to be the best initial intervention. However, if the student does not respond to the email, a phone call follow–up became the most effective approach.

Final Discussion and the Road Ahead

Each of these case studies involved institutions doing the work of developing deep insight analytics capacity and deploying action analytics strategies. From the results of these and other projects across our institutional cohorts, we point to the following observations as keys to leveraging these strategies to impact student learning and completion work:

Insight analytics that are developed using institution–specific data sources – particularly student–level SIS and LMS data – are vital to understanding student flow, as well as targeting and personalizing intervention and outreach. In short, there is not a global predictive model that works across institutions with any level of accuracy. You need to "turn the lights on" in your institution.

- The inclusion of additional data streams in insight analytics work can add value in better understanding student flow and targeting outreach.
- Adding ongoing activity data from students improves the performance of model predictive power.
- Bringing insight analytics together with action analytics is essential to "moving the needle" on student success. Better precision of the models helps target outreach and improve impact of instruction and advising support.
- Trying and testing action analytic outreach is a must. The work of iterating on outreach, what some in our community are calling intervention science, results in the best outcomes. There are no silver bullets, and tuning outreach to a unique student population is key. Put simply, the predictive models are just the beginning of the work.
- How you bring data to the front lines of learning – e.g., to advisors and faculty – has a significant impact on the effectiveness of these efforts. Modality, timing, visualization, and operational tools matter.

We summarize these findings in a simple framework we call the challenge of the four rights: (a) building the right infrastructure to (b) bring the right data to (c) the right people in (d) the right way. Importantly, the right way may be the most difficult aspect, because it includes how we visualize data, operationalize interventions and outreach, choose modalities, provide real–time feedback, and test the timing of interventions and outreach. In many ways, this is the art and science of analytics initiatives in higher education. Moreover, we need to ensure that we take security, privacy, and especially the impact of unintended consequences seriously. Indeed, data brought the wrong way to at–risk students – e.g., a flashing red indicator that in essence tells them that they are destined to fail – might do great damage to a population we care about a great deal (Stevens, 2014). That is why the trying and testing of outreach as a discipline is key here.

Moreover, we need to ensure that we take security, privacy, and especially the impact of unintended consequences seriously. Indeed, data brought the wrong way to at–risk students – e.g., a flashing red indicator that in essence tells them that they are destined to fail – might do great damage to a population we care about a great deal.

Going forward, the work of the Civitas Learning community will be focused on how we continue to bring together the best of insight and action analytics to help students learn well and finish strong on higher education pathways. Much is to be done, and much is to be learned. But as the field of analytics continues to take shape in higher education, there is clearly great promise. However, learning together will be essential.

References

Blumenstyk, G. (2014, September 15). Companies promise "personalized education." *The Chronicle of Higher Education.* Retrieved from http://chronicle.com/article/Companies-Promise/148725/

Christakis, N., & Fowler, J. (2007). The spread of obesity in a large social network over 32 years. *The New England Journal of Medicine, 357*(4), 370–379.

Civitas Learning Space. (2014). Retrieved from http://www.civitaslearningspace.com/

Complete College America. (2014). Retrieved from http://completecollege.org/

Crosta, P., & Kopko, E. (2014). Should community college students earn an associate degree before transferring to a four-year institution? *CCRC Working Paper No. 70.*

Deeper Learning Initiative. (2014). *The William and Flora Hewlett Foundation*. Retrieved from http://www.hewlett.org/programs/education/deeper-learning

Education Initiative. (2014). *The Kresge Foundation*. Retrieved from http://kresge.org/programs/education

Fain, P. (2014, March 27). Predicting success. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/news/2014/03/27/civitas-brings-completion-oriented-big-data-community-colleges-and-universities.

Fisher, R. A. (1954). *Statistical methods for research workers.* Edinburgh, UK: Oliver and Boyd.

Higher education. (2014). *Obama administration 2020 goals.* Retrieved from http://www.whitehouse.gov/issues/education/higher-education

Hill, J., & Su, Y.-S. (2013). Assessing lack of common support in causal inference using Bayesian nonparametrics: Implications for evaluating the effect of breastfeeding on children's cognitive outcomes. *The Annals of Applied Statistics, 7*(3), 1386-1420.

Kim, J. (2014, January 26). 8 questions for the co-founders of Civitas Learning. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/blogs/technology-and-learning/8-questions-co-founders-civitas-learning

Lumina Foundation strategic plan 2013-2016. (2013). *Lumina Foundation*. Retrieved from http://www.luminafoundation.org/goal_2025.html

Milliron, M. (2013, December 9). Confusing signals. *Inside Higher Ed*. Retrieved from https://www.insidehighered.com/views/2013/12/09/essay-real-meaning-debate-over-early-notification-system-students

Milliron, M. D., & Rhodes, R. M. (2014). What do we expect? A regional education ecosystem that works. *The Texas Lyceum Journal*, 15-22.

Phan, N., Xiao, X., Dou, D., Piniewski, B., & Kil, D. (2014). Analysis of physical activity propagation in a health social network. *ACM International Conference on Information and Knowledge Management*, Shanghai, China.

Postsecondary success strategy overview. (2014). *Bill and Melinda Gates Foundation*. Retrieved from http://www.gatesfoundation.org/What-We-Do/US-Program/Postsecondary-Success

Rees, J. (2014, September 3). September Pioneer Summit 2014. *Civitas Learning Space*. Retrieved from http://www.civitaslearningspace.com/september-pioneer-summit-2014/

Stevens, M. L. (2014). An ethically ambitious higher education data science. *Research & Practice in Assessment, 9*(2), 96-98.

Texas Student Success Council. (2014). *Educate Texas*. Retrieved from http://www.edtx.org/postsecondary-access-and-success/postsecondary-success/texas-student-success-council/

Thornburgh, C., & Milliron, M. D. (2013, December). Driving analytics to the front lines of education. *CIO Review*. Retrieved from http://www.cioreview.com/magazine/Driving-Analytics-to-the-Front-Lines-of-Education-FKCU497192190.html

Appendix

Deep Dive on Some of the Data Science behind Insight and Action Analytics

Overview of Insight and Action Analytics

Extracting actionable insights from data requires a complementary fusion of (a) extraction of insightful derived features, (b) ranking and optimization of features in a hierarchical learning network to accommodate a diverse collection of data footprints of students, and (c) visual analytics to surface complex information in an intuitive way.

Feature extraction is a continuous quest to encapsulate and bring to light useful information that can be acted upon. In this Appendix, we show examples of various insights in one-, two-, and multi-dimensional plots in an increasing order of complexity. Figure 1 shows a few examples of insightful features in marginal class-conditional densities. The probability density functions (PDFs) in green and orange are $p(x/y$=persist) and $p(x/y$=not persist), respectively, where x = student feature and y = student success outcomes or classes in classification.

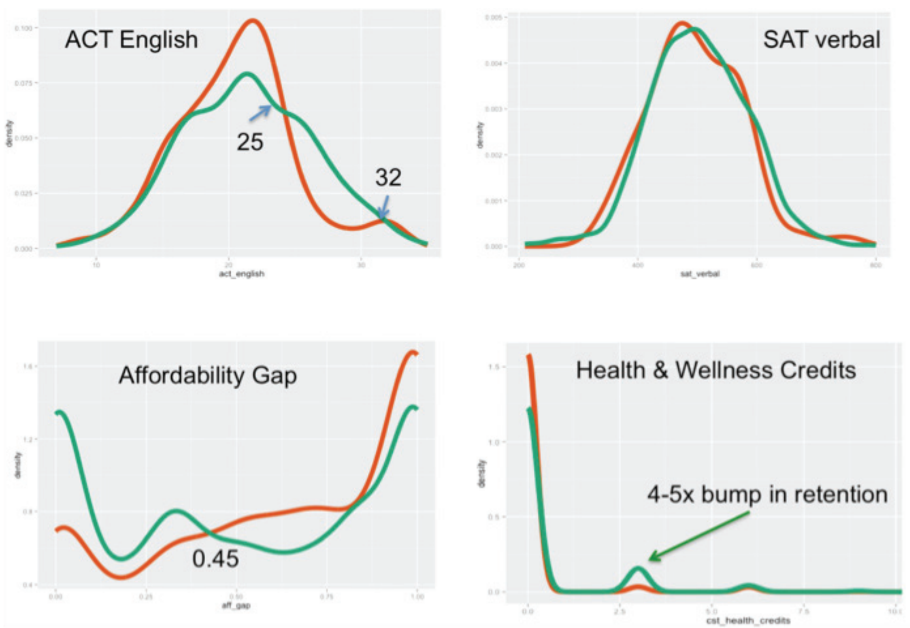


Figure 1. Examples of insightful features. With the exception of plot from the ISSM model, the rest are derived from persistence prediction models.

The ACT English plot is interesting in that SAT Verbal was not a strong predictor of persistence. When we probed deeper, we learned that this institution places a heavy emphasis on writing in all their courses. ACT English measures writing skills while SAT Verbal does not.

Another example is that the affordability gap (AG) shown in the lower left-hand corner is more insightful than raw financial aid amount since AG measures the ratio of financial aid to tuition owed. Such a plot can provide insights into how to allocate Pell Grant financial aid to improve persistence of Pell Grant recipients.

The Health & Wellness plot shows that students who take one health & wellness course as an elective persist at a much higher rate. While this observation does not imply causation, it can lead to an interesting research question and experiment design

The class-conditional feature PDFs compare incoming student success rates as a function of the percentage of single-parent households in zip codes students come from. An actionable implication here is that if an incoming student has a high risk of not persisting and is from a high single-parent household area, she may be a prime candidate for a mentorship program, especially if a mentor has a similar background in the beginning, but has been academically successful with good social skills.

In certain situations, a combination of more than one feature brings out more meaningful insights. Pathway analysis has generated a lot of interest, especially for community college (CC) students (Crosta & Kopko, 2014). Figure 2 shows clearly that the probability of earning a bachelor's degree reaches a peak at around 60 credit hours. That is, CC students who earn AA/AS degrees improve their probability of earning bachelor's degree by more than 10% from the baseline trend for all transfer students.

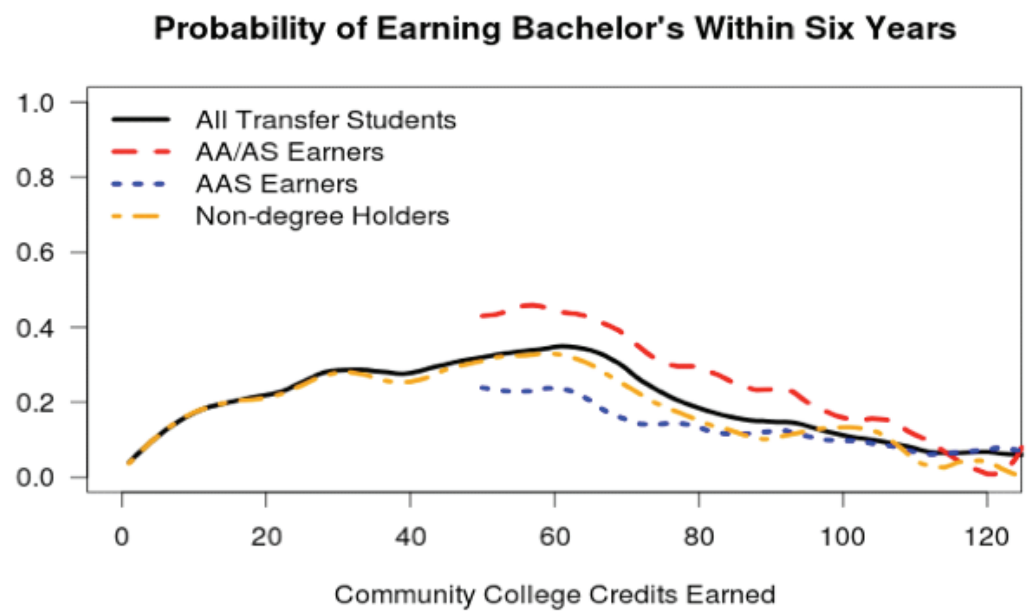


Figure 2. College pathway analysis (Crosta & Kopko, 2014).

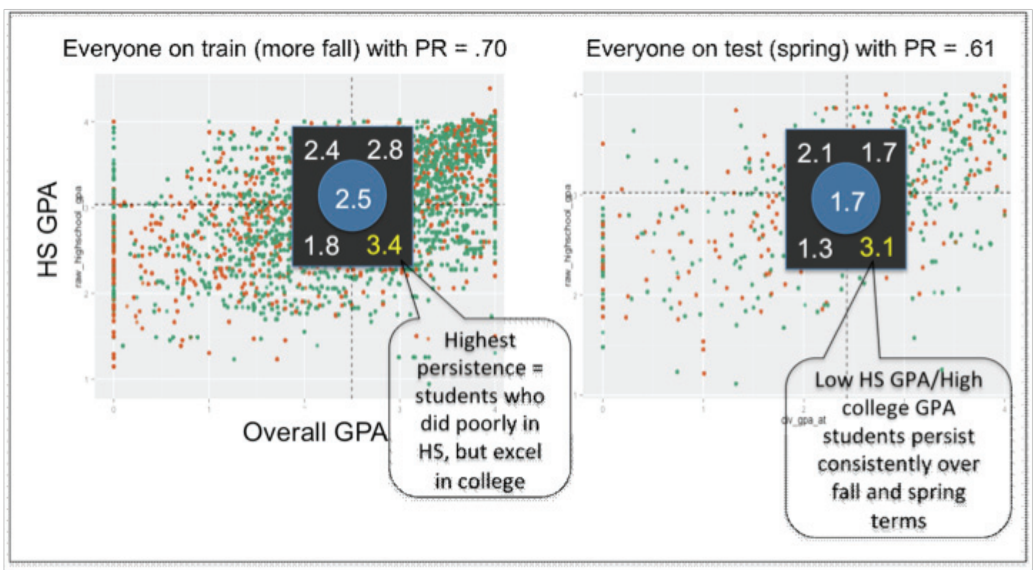


Figure 3. The 2 x 2 scatter plots over high school GPA and community college GPA paint an interesting picture. The five numbers in the centroid (50%-50% line) represent the ratio of the number of students who persist to that of students who do not for all and each of the four quadrants. Persistence rate drops significantly in spring, in part due to high-performing students transferring out.

We are currently federating data between 2- and 4-year schools, where the 2-year institutions serve as feeder schools to the 4-year institutions, so that we can do more thorough investigation into optimal transfer pathways and how to apply personalized interventions to students who are likely to benefit by finishing AA/AS degrees at community colleges.

In general, students with high CC GPA in the spring term tend to transfer out, which may suggest that advisors should target high-GPA students in the spring term to help them be better prepared by staying an extra year to earn AA/AS degrees. However, when we overlay another feature, high school GPA, a more interesting picture emerges as shown in Figure 3.

The 2 x 2 scatter plots use the same color code as in Figure 1. Each scatter point represents a student with color denoting the persistence flag (orange = not persist, green = persist). The number in the blue circle represents the ratio of those who persisted to those who did not. The four numbers along the edge depicts the same numbers in the four quadrants along the centroid.

The first observation is that the persistence rate (PR) is much lower in spring. The second key finding is that students with low high school GPA and high CC GPA (quadrant 4) tend to persist at a much higher rate than those with high GPAs in high school and CC, as well as their persistence rate being less dependent on seasonality. This finding alone can help advisors improve their targeting. Another example deals with the impact of scholarship on persistence as shown in Figure 4.

The left plot shows that merit scholarships given to students with high ACT scores are not as effective as those given to students with high high-school GPA. What is also interesting is that students who have high school GPA tend to persist at a higher rate than those with ACT scores. This shows the importance of multidimensional decision making by factoring into all key drivers of student success that depend on which segments and clusters they belong to in the hierarchical learning network based on data availability and clustering within each data-availability segment.

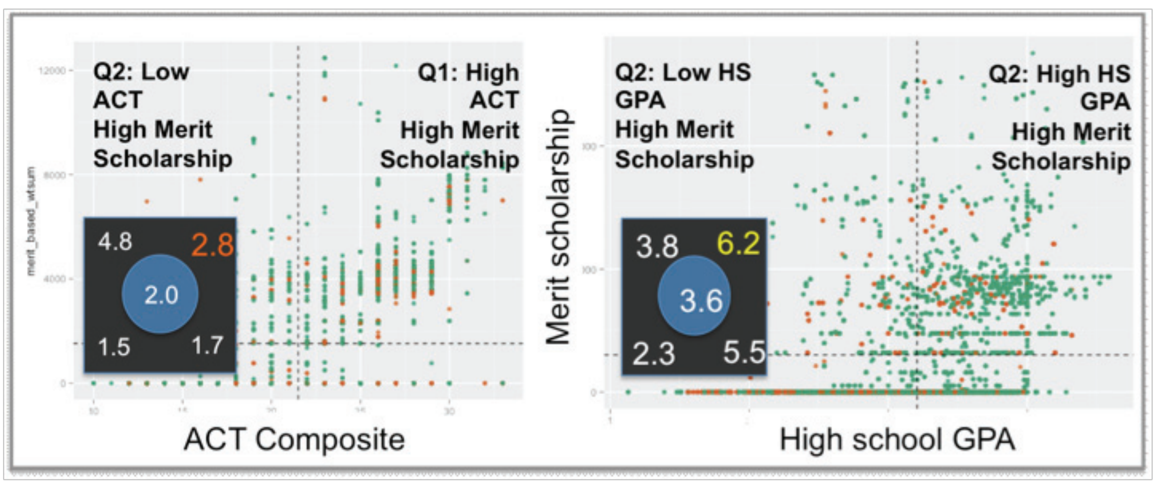


Figure 4. The impacts of scholarship on persistence.

Now we can extend the 2 x 2 concept indefinitely to provide insights with an arbitrary number of top features and/or at the segment/cluster level, where segments are determined based on available data footprints while clustering finds homogeneous groups within each segment, thus facilitating a hierarchical network view of the entire student population. Figure 5 shows the cluster heat map view. Columns and rows represent clusters and z scores (mean/standard deviation) of various attributes that characterize each cluster. The first two rows are population size (N) and persistence rate of each cluster. The rest of the rows represent various attributes, such as census household income, % of population with BA degree or higher based on census, age, cumulative GPA, the number of distinct 2-digit CIP codes in course work per term, etc. This quilt view extends much further in reality, while highlighting differences among the clusters based on color gradients across each row. Figure 5 shows 3 sets of clusters (low, medium, and high) grouped based on actual persistence rates. Table 1 compares and contrasts these performance-based clusters.

Furthermore, graph theories can be applied to understanding course pathways and the impacts of emerging influencers and cliques on helping other students succeed. Figure 6 shows a concurrent social graph and a time-varying series of student social networks over the course of a term.

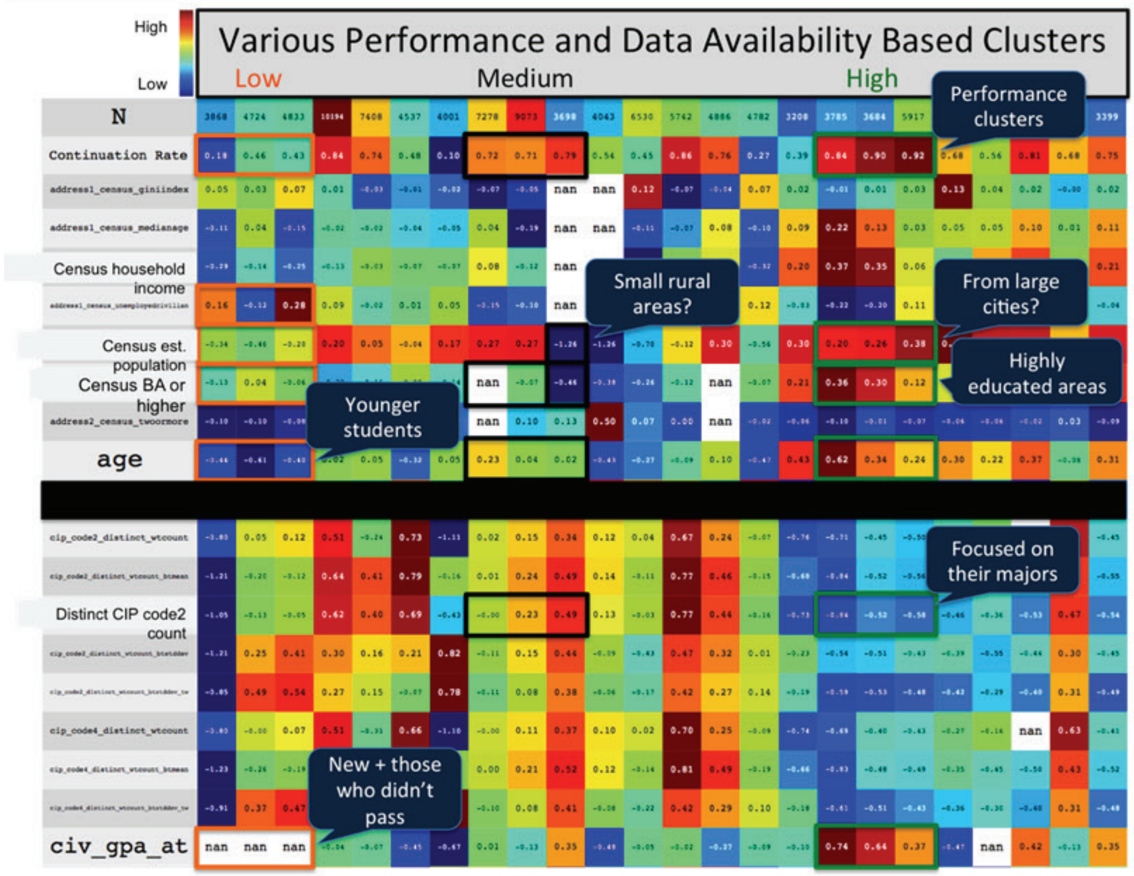


Figure 5. The cluster heap map view so that we can glean insights into how these clusters can be differentiated based on demographic variables, census-derived features, and top predictors. The white color indicates that the corresponding features and their associated raw data elements do not exist.

Table 1

A Comparison of the Three Performance-Based Clusters

Attribute	Low persistence	Medium persistence	High persistence
Persistence rate	~40%	~75%	~92%
Census household income	Low	Medium	High
Census estimated population	Small	Mixed	Large
% of residents with BA or higher degrees	Low – med	Low	High
Student age	Young	Middle	Mature
Distinct CIP code2 count	Low to med	High	Low
Cumulative GPA	N/A	Medium	High
Financial aid	Pell + some loan	Some Pell, little loan	High loan, little Pell
Terms completed	None	Most	Middle

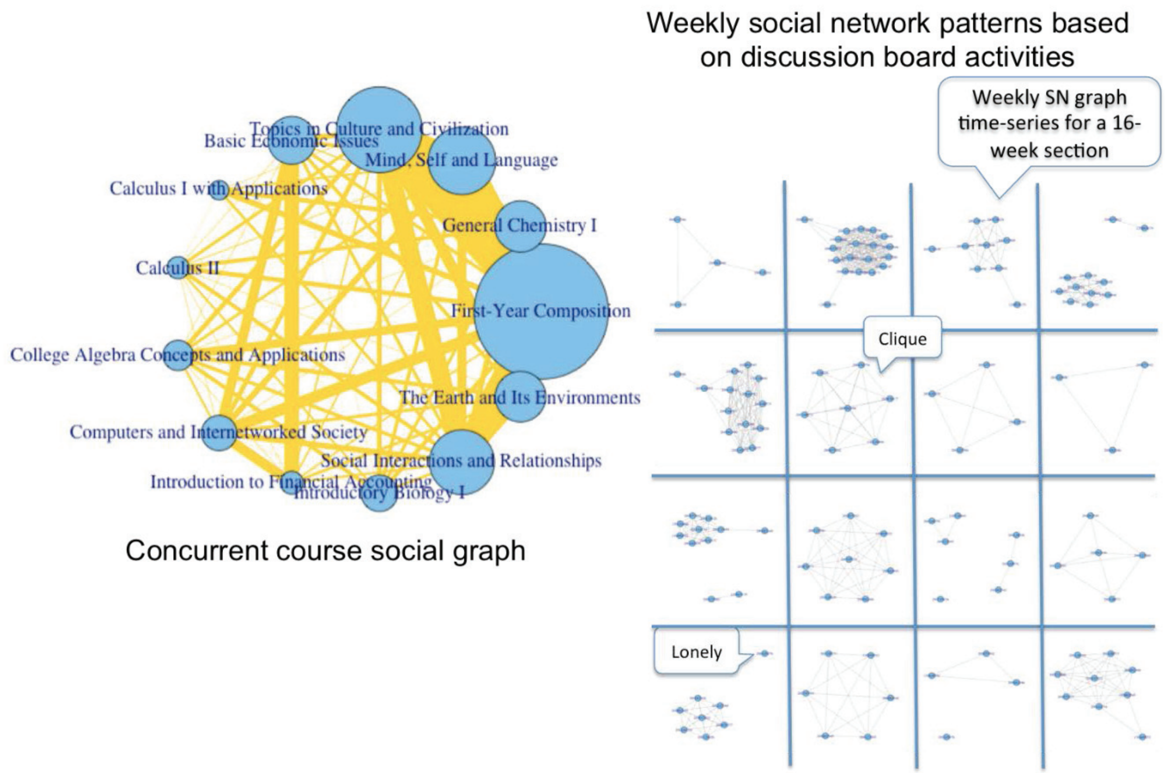


Figure 6. Course social graph and social network dynamics throughout a term.

The concurrent course social graph shows what courses are being taken together with the vertex size proportional to enrollment. The thickness of edges between courses is proportional to how frequently the connected courses are taken together. This allows us to investigate students' course-taking behaviors and toxic/synergistic course combinations by melding successful course completion predictions, propensity score matching by creating test and matching control groups, and explicitly incorporating students' course-taking patterns. The same analysis can be extended to course pathways over multiple terms to help us glean insights into the paths taken by successful vs. less successful students.

The same concept applies to social network analysis. Christakis and Fowler (2007) found that obesity spread through one's social network. Phan et al. (2014) applied the concept further by identifying emerging influencers and then studying their influence on connected pilot participants as a function of time to quantify how good health behaviors can be spread through peer-to-peer nudging, discussion board, and sharing of pedometer data through games. We plan to apply similar methodologies in student social networks so that we can work with faculty in facilitating students helping other students under faculty nudging. Our preliminary work indicates that a few social network features are statistically significant in predicting successful course completion and persistence.

Examples of Action Analytics

Action analytics can be most effective when actionable insights are brought to frontline people and their intervention details are captured in database tables for an integrated predictive and intervention science research. In principle, the predictive science provides insights into who is at risk, when the right moment for engagement or intervention is, and what intervention will be effective down to an individual student level. Intervention science works in concert with predictive science to provide foundational data for computing intervention utility, which in turn becomes the basis for intervention recommendation.

Intervention science data comes from encoding all facets of interventions – type, delivery modality, messaging attributes, business rules for intervention (who, when, and why), and primary/secondary endpoints for outcomes. Intervention science analytics encompass experiment design, power analysis, propensity score matching (PSM), Bayesian additive regression trees (Hill & Su, 2013), predictive modeling, and predictive ratio analysis. All these methods can shed scientifically rigorous insights into what interventions work or do not for which groups of students under what context. Figure 7 shows our overall framework for intervention science.

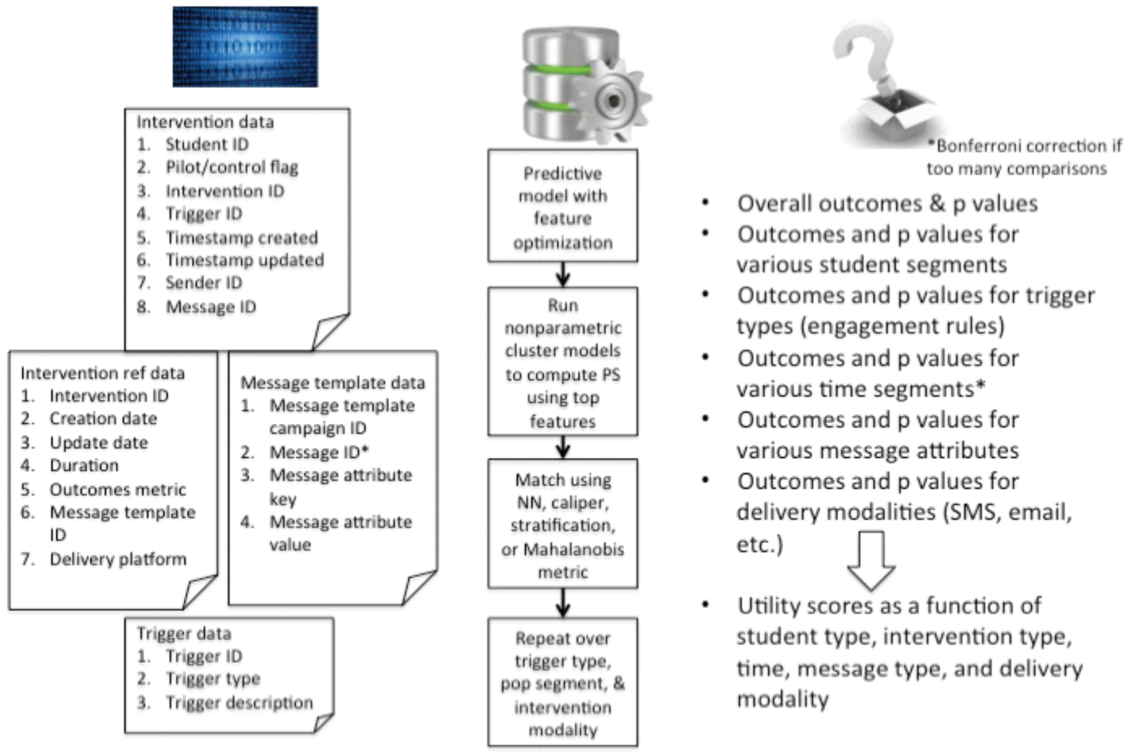


Figure 7. Our intervention science framework that leverages both predictive models and drill-down outcoes analytics to provide insights into intervention efficacy.

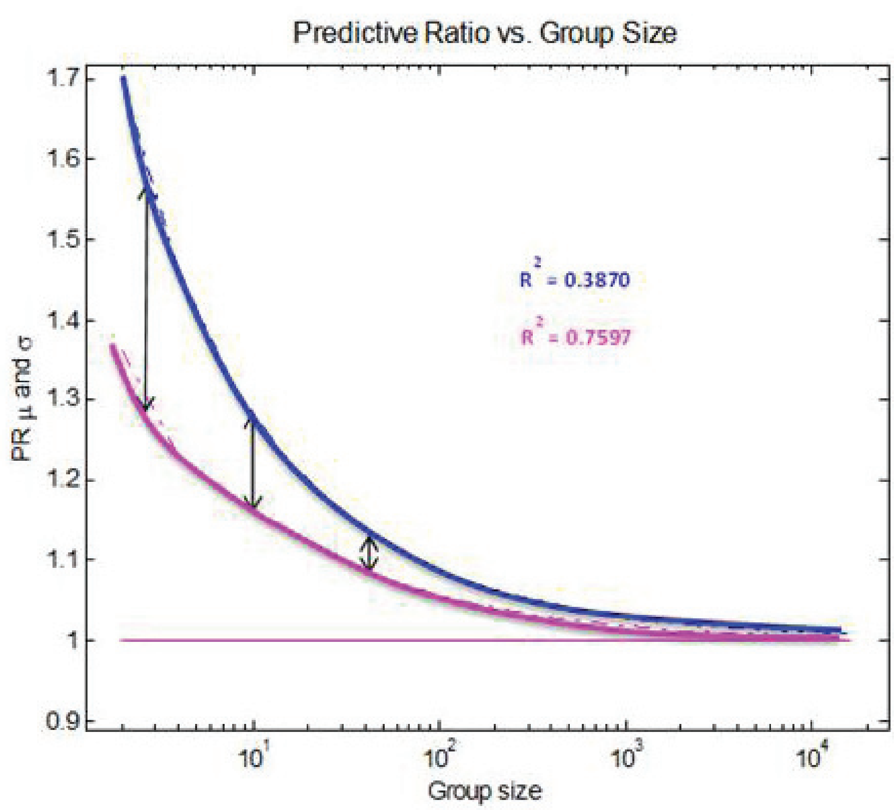


Figure 8. The more powerful the model is measured by R2, the smaller the standard deviation in predicitve ration (PR) is at various group sized, leading to greater statistical power, i.e., a lower minimum detectable threshold in outcomes differences between pilot and control.

Action analytics apps surface to frontline prediction scores and key risk drivers at an individual student level. They also provide real-time feedback on intervention efficacy by showing how student engagement scores, prediction scores, and early enrollment statistics are changing for the pilot group in comparison to the control group. We select students in the control group through randomization and/or PSM prior to the commencement of a pilot.

In order to maximize statistical power in outcomes analysis, we apply hierarchical modeling techniques based on data availability, where a model is instantiated at the segment level. For each segment, we use the model's top features in PSM. The more predictive the models are using these top features, the greater the statistical power is. Figure 8 demonstrates that the higher-performance model in magenta exhibits a lower standard deviation curve for predictive ratio at all group sizes. Furthermore, we augment PSM with prediction-score matching such that matching cohorts have similar PDFs in propensity and prediction scores.

In summary, action analytics take risk predictions as an input in order to identify when to apply which interventions to which students. Once interventions are applied, we use various primary and secondary endpoints to investigate the efficacy of interventions as a function of engagement business rules, population segments, and intervention modalities. We provide real-time feedback for advisors and faculty by pointing out how their interventions are affecting feature and prediction score PDFs since human factors also play such an important role in affecting intervention outcomes. This information becomes the foundation of action analytics and intervention science.

Book Review

Uncharted: Big Data as a Lens on Human Culture.
Erez Aiden and Jean–Baptiste Michel. New York, NY:
Riverhead Books, 2013. 288pp.
ISBN–13: 978–1594487453. Hardcover, $19.16.

REVIEWED BY:
Carolyn Penstein Rose, Ph.D.
Carnegie Mellon University

As pressures to scale up education and assessment mount higher and higher, attention has turned towards techniques from the field of big data analytics to provide the needed boon. At first blush, Aiden and Michel's book *Uncharted: Big Data as a Lens on Human Culture* would not seem to speak to this issue directly, yet it does provide the opportunity for some needed reflection.

As pressures to scale up education and assessment mount higher and higher, attention has turned towards techniques from the field of big data analytics to provide the needed boon.

The vision of the idealized data science of the future has recently been characterized as something akin to archeology and geology (Knight et al., 2014), two fields where scientists conduct painstaking, careful, and reflective work to reconstruct the past from the fragments that remain. This characterization of our work challenges us to take greater care as we piece together evidence of psychological and social processes from the digital remains of cognitive and social activity taking place within the online world. In particular, it challenges us to take a step beyond just counting what can be easily counted, and push for greater theoretical depth and validity in our attempts at quantification and operationalization as we seek to make sense of the signals we can uncover using the growing number of powerful modeling technologies that have been developed in recent decades.

Within this sphere, Aiden and Michel's book is a popular press treatise designed to introduce a nontechnical readership to the capabilities of the Google Ngram Viewer.[1] It presents a fascinating new look at history through the lens of "robots," which are automated lexicographers that index arbitrary lengthed word sequences, referred to as ngrams, as they occur within the expanding Google Book collection.[2] The ngram viewer makes its debut in the book by producing a graph that challenges a claim about the historical event that triggered a shift in how the "United States" is treated grammatically, i.e., whether we treat it as a plural reference to a multiplicity of states or a singular reference to a collective whole. The shift in grammatical status is purported to reflect a shift in conception, and therefore has great historical significance, especially to Americans. The evidence of such a shift in usage is a graph of relative frequency of occurrence of "The United States are" and "The United States is" over time in the Google Book collection. The shape of the displayed trend is different from what one might think if it did indeed reflect the change in conceptual status and was indeed triggered by a historical event in that, it occurred gradually rather than suddenly, and it was not until fifteen years after the event that was believed to have triggered it when the dramatic difference in preference emerged. The reader is challenged to consider the extent to which previous conceptions of history might be challenged by viewing it through the eyes of these robot lexicographers.

The Google Ngram Viewer is a text visualization tool (Siirtola, Saily, Nevalainen, & Railha, 2014). One can consider its representation of text as something of a cross between world clouds, which give a cross–sectional view of word distributions from a corpus in graphical form, and graphs of topic trends, which use dimensionality reduction techniques like Latent Dirichlet Allocation (Blei, Ng, & Jordan, 2003) or Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998) to identify themes and then plot the relative prevalence of those themes over time within a corpus. Word clouds are often used to suggest the values communicated through a text or text collection by displaying words with a relative size that indicates their relative frequency, with the implication that relative frequency says something about relative value. Topic trends present a more digested view, in that they collapse together sets of words that co–occur, and therefore might function together as elements that together communicate a theme. The representation of these automatically identified themes as a graph of their relative frequency over time is displayed through line graphs arguably provides a much coarser grained perspective on what is in the text, and yet it offers the possibility of comparing topic focus between different periods of time. And its coarser grained representation better leverages the richness in stylistic variation that language affords. Like a word cloud, the Google Ngram Viewer's representation displays relative frequency of ngrams as a representation of relative value. But unlike the cross–sectional nature of a word cloud, its representation allows us to see trends over time. Similarly unlike word clouds, it is extremely selective in which relative frequencies it displays. Thus, unlike topic trend representations, it does not consider the great variation that language affords in referring to an idea, or even in realization of a specific lexical construction. A rigorous interpretation of the significance of the graphs would take these contrasts into account.

The first chapter of the book recounts the history of the development of the Google Ngram Viewer and illustrates its use with some key examples. After that, with each of the next five chapters, a new and fascinating question that might be investigated using this tool is introduced and explored. The Google Ngram Viewer is posed as the data analyst's correlate of Galileo's telescope. While the richness of the signal provided by such a viewer is admittedly impoverished, it is compared to the remnants of monetary systems of old left behind for anthropologists to use to piece together an

[1] https://books.google.com/ngrams
[2] http://books.google.com/

image of cultural practices of old. The authors pose questions about the status of theory in light of the great multitude of hypotheses that can be imagined and quickly tested with such a resource.

While the authors compare the Google Ngram Viewer to the telescope of Galileo, the book does not come across to my academic ears as designed as a serious foray into data science, nor meant to make serious contributions to the fields of humanities and social sciences. To its credit, it raises some methodological concerns even in the first chapter where the authors affirm the need to validate interpretations from quantifications and acknowledge the difficulty of doing so in a corpus as large as the Google Books archive. Thus, it would not be fair to critique it based on methodological standards of the fields of data science. Nevertheless, it is useful in the context of a special issue on learning analytics, and assessment specifically, to consider what message this book might have for us as a community as we reflect on our own practices of scientific inquiry.

Nevertheless, it is useful in the context of a special issue on learning analytics, and assessment specifically, to consider what message this book might have for us as a community as we reflect on our own practices of scientific inquiry.

Consider the following anecdote. A recent New York Magazine article reported that personnel at Pinterest had noticed a strong trend for numerous women to collect substantial numbers of pins related to weddings. The interpretation of this strong focus on weddings was that these women were most likely preparing for their respective weddings. Thus, the organization then proceeded to send an email to them with text that implied they were indeed preparing to get married. It turned out, however, that most of them were single and were collecting the pins for other reasons. Some responded in a way that suggested they were dismayed at the mistake. This anecdote illustrates well how easy it is to misinterpret what a pattern might be telling us, even when the pattern appears strong and clear. The problem is that Pinterest was not designed to provide others with insight into the reasons why people are interested in or collect the items that they do, and thus it is not valid to assume that upon viewing ones pins the viewer would get insight into these reasons.

Similarly, in the case of the Google Ngram Viewer, it is easy to imagine that while the view provided by the robots has some advantages over our own human perspective on history (e.g., perfect memory, long time view, ability to consider every word in the entire book collection, etc.), we must consider the important ways in which the view it provides might be obscured by what its missing. For example, the contrast between "The United States is" and "The United States are" neglects the fact that the great majority of mentions of the phrase do not place it as the subject of the copula, and therefore will be skipped in this analysis.

Furthermore, the contexts in which it is positioned this way are not a random sampling of mentions since this form is indicative of a definitional statement, although the grammatical treatment of the phrase in other contexts is equally a reflection of the conception of its status as an entity. It is equally important to note that books included in Google Books might not be a random sampling of published books, and the language of book publications might not be a random sampling of language produced. Furthermore, the analysis fails to take into consideration that many genres of writing include language that reflects not the style or perspective of the author, but perhaps the style or perspective of a synthetic culture created as a fictional character or culture, or the author's potentially mistaken conception of how some other would present him or herself. All of these issues and more threaten the validity of the conclusions one might draw from the graphs, no matter how compelling they might appear.

Coming back to the focus of this special issue, what does this tell us about the use of big data analytics for assessment? The book is well worth a thoughtful read by all who look to big data analytics to play a growing role in large scale assessment. It is not to say that the book should either encourage or discourage such a movement. It should simply provide the opportunity to reflect on issues related to validation of interpretation. And specifically with respect to assessment based on analysis of textual data, issues related to the incredible richness and variability of language usage should be appreciated and allowed to raise an appropriate level of skepticism.

References

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

Knight, S., Wise, A., Arastoopour, G., Shaffer, D., Shum, S. B., Kirschner, P., & Collier, W. (2014). *Learning analytics: Process vs practice*. Presentation at the at Learning Analytics for Learning and Becoming in Practice Workshop, International Conference of the Learning Sciences, Boulder, CO.

Landauer, T., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis *Discourse Processes, Special Issue: Quantitative Approaches to Semantic Knowledge Representations, 25*(2–3), 259–284.

Roy, J. (2014, September 4). Pinterest accidentally congratulates single women on getting married. *New York Magazine*. Retrieved from http://nymag.com/daily/intelligencer/2014/09/pinterest–congratulates–single–women–on–marriage.html

Siirtola, H., Saily, T., Nevalainen, T., & Railha, K.–J. (2014). Text Variation Explorer: Towards interactive visualization tools for corpus linguistics. *International Journal of Corpus Linguistics, 19*(3), 417–429.

Book Review

Building a Smarter University: Big Data, Innovation, and Analytics. Jason E. Lane (Ed.). Albany, NY: State University of New York Press, 2014. 325 pp. ISBN–13: 978–1438454528. Hardcover, $81.00. Paperback, $27.95.

REVIEWED BY:
Fabio Rojas, Ph.D.
Indiana University – Bloomington

The dam has broken. We are now awash in a deluge of data so large that it has its own special name, "big data." This is not a bad thing, nor is it totally unexpected. Sooner or later, social scientists and policy makers were going to get their hands on the data that people generate as they use the Internet. Already, such data have helped researchers understand political trends, health seeking behavior, and economic fluctuations. Now, it is time for higher education researchers to face the challenge of big data. What is big data in higher education? How can it be used? A new book, *Building a Smarter University: Big Data, Innovation, and Analytics*, tries to answer these questions with a series of essays written by higher education professionals.

Roughly speaking, innovations trigger three types of responses. First, people ask "What is this?" Second, one may ask, "What can we do with this?" And third, one may ask, "What are the rules for doing this?" *Building a Smarter University* has chapters addressing each question.

Roughly speaking, innovations trigger three types of responses. First, people ask "What is this?" Second, one may ask, "What can we do with this?" And third, one may ask, "What are the rules for doing this?"

When innovations emerge, practitioners try to make sense of the new phenomenon. They did not learn about the new technology in graduate school and that raises unexpected issues. Early in the history of a technology, one will encounter essays that focus on definitions, examples, and guidelines for practice. One might call this the exegetical phase of a new science. At this point, scholarship is more about sense–making than problem oriented "normal science." It is about explaining things to a puzzled audience. At times, this can be productive. People need definitions, a key to help them understand what is new and why it deserves attention.

People need definitions, a key to help them understand what is new and why it deserves attention.

Building a Smarter University has its fair share of explanatory essays, such as Lane and Finsel's chapter that explains the "basics" of big data and why people might care. Some readers might be familiar with the basic themes, but the basics of big data bear repeating. Basically, big data is usually characterized by its size, speed, and continual creation. There is an emerging definition codifying this idea: big data has "five V's": Volume, velocity, variety, veracity, and value. While I do not dispute this basic intuition, it often misses something important. Big data is native to the Internet and the computing world in ways that older types of data are not. It is also natural in the sense that it was not concocted by a researcher in a survey or interview.

This is an important distinction for higher education researchers. For example, consider the typical student affairs professional who now has access to real time data on how students search for classes from their institution's online catalog. While size and speed may be important, the crucial issue is that this is a more accurate reflection of a student's shopping behavior than what people report in surveys or focus groups. Similarly, if one were interested in bolstering minority enrollment, it might be better to monitor social networks than rely on self–reports of the college experience. The reason is that the Internet sometimes encourages a more candid discussion of issues than the manufactured environment of the focus group or survey. The Internet also records real behaviors as well. That is the true value of big data, not necessarily its speed or size.

Big Data is usually characterized by its size, speed, and continual creation.

While *Building a Smarter University* has some fine exegetical chapters, there are some that are less helpful because they use big data to pursue philosophical points that typical practitioners will not find relevant. For example, Bringsjord and Bringsjord use big data to illustrate a theory of information ("big data" vs. "big–but–buried data") and relate it to Zeno's paradox. There is a valid point to be made that raw information and knowledge are different things, but I am not sure that such an esoteric presentation is helpful. Even though I took courses in mathematical logic in college, I honestly found it difficult to relate their approach to what the typical higher education researcher would find helpful.

Once people know about innovation, the question becomes application. People want a sense of how a new resource can be used to solve specific problems. It is here that *Building a Smarter University* has the most to offer. Numerous chapters offer concrete examples of how this new type of data can help administrators make colleges better. Indeed, given how difficult it is to change or affect student behavior, it is refreshing to see creative applications of big data.

Ben Wildavsky's chapter is one excellent example of an application of big data to student affairs. Normally, student affairs professionals must react to student performance. A student may meet an advisor after they have received a bad grade, or are at risk of failing the course. Often, an advisor can not help the student because their current score is so low that even an exceptional performance in the rest of the course will not save them. Instead, what if the advisor had real time access to the student's performance? Or models that would project grades based on the performances of thousands of earlier students? Perhaps, there might be a real time warning system. As the instructor enters grades, students with poor performance might have a warning signal sent to an advisor.

Big Data is native to the Internet and the computing world in ways that older types of data are not.

Such a system that continually monitors, tracks, and assists students with course selection would be enormously useful (Denley, 2014; Milliron, Malcolm, & Kil, 2014). It would be a vast improvement over the current system where advisers go on a high school transcript and good intentions. In some cases, they rely on second hand knowledge of courses handed down by earlier generations of students. Considering that a college degree carries an enormous premium on the labor market, helping a student complete their degree using advice derived from a big data model could be of enormous importance.

Other chapters by Goff and Shaffer, Owens and Knox, and Lane and Bhandari touch on financial aid, identifying course equivalencies, and measuring the globalization of higher education. It is not too hard to imagine that organizational strategy in higher education would be impacted by big data. Enrollments and recruitment could be measured, faculty productivity monitored, and fund raising can be optimized.

There is the question of ethical and legal standards. *Building a Smarter University* has a number of chapters that address the legal aspects of big data. Jeffrey Sun's chapter is a nice review of the relevant privacy issues. The primary issue is how FERPA applies to student generated data. In general, such data can be used internally for research purposes, but complexities arise when a university has branches that are located outside the United States, or in states where privacy rules differ. As administrators try to use this data, there will be an effort to provide some clarity and uniformity on these issues.

While there have been earlier attempts at harnessing college generated data, we simply have not had the tools to effectively use that information.

This book shows how big data can be an important tool for higher education administrators. While there have been earlier attempts at harnessing college generated data, we simply have not had the tools to effectively use that information. *Building a Better University* shows how that might change.

References

Denley, T. (2014). How predictive analytics and choice architecture can improve student success. *Research & Practice in Assessment,* 9(2), 61-69.

Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. *Research & Practice in Assessment,* 9(2), 70-89.

Book Review

Assessing the Educational Data Movement.
Philip Piety. New York, NY: Teachers College Press, 2013. 223 pp.
ISBN–13: 978–0807754269. Paperback, $35.10.

REVIEWED BY:
Karly Sarita Ford, Ph.D.
Penn State University

Although Philip Piety's book, *Assessing the Educational Data Movement*, is written about the educational data movement in the K–12 sector, it provides many novel ideas and cautionary tales for researchers and practitioners of higher education assessment.

While we are all aware of the technical part of educational data, its social and revolutionary impacts are not to be discounted.

Piety frames the book by suggesting that educational data movement is a *sociotechnical revolution*. While we are all aware of the technical part of educational data, its social and revolutionary impacts are not to be discounted. Like the telegraph or the cell phone, the development of educational data has shaped our social lives, the way we think, interact and live. Thinking about educational data as a technical development with wide ranging social impacts immediately turns a narrow subject into a roaming intellectual landscape. Suddenly, we are not only examining math test scores of third graders; we are able to think about how teachers respond to pressures, how schools shift schedules to accommodate testing, how parents consume school report card data and how district budgets are rewritten to include teams of educational data scientists (Macfadyen, Dawson, Pardo, & Gasevic, 2014). It is that we now have a job title "educational data scientist." Indeed, the educational data movement has deep social impacts and naming it as a sociotechnical revolution is Piety's first intellectual gift to his readers.

Indeed, the educational data movement has deep social impacts and naming it as a sociotechnical revolution is Piety's first intellectual gift to his readers.

The introductory sections of the book provide a brief history of the US Department of Education's shift toward data use. Piety describes the historical context for the introduction of the Institute of Education Sciences (IES) in 2002. At the time the agency was entirely focused on randomized control trials (RCTs). In recent years we have seen IES move away from RCTs and fund projects with a range of methodologies. Another turning point in the data movement was the introduction of No Child Left Behind (NCLB). The central indicator was Adequate Yearly Progress (AYP), a school level measure that proved to have many problems, perhaps the worst of which was the assumption that the population of school did not vary much from year. Having learned from the pitfalls of AYP, the in vogue assessment strategy are Value Added Models, which focus on individual improvement from one year to the next.

Piety convincingly argues that education and business, two communities that are often painted as being culturally and substantively separate, are more conceptually linked than we might think. This of course is a minefield, where many education researchers and practitioners balk at education being viewed as a process that could be compared to automated efficiency and bottom line driven private sector. However, Piety traces how the world of business first reacted to and integrated data into its own operations. While customer service and executive resource planning were once siloed parts of the corporate enterprise, data collection and analysis connected them – requiring them to communicate with more regularity and creating less rigid boundaries between sectors. The parallel example in the education world would be how data has linked district level offices to classrooms. Where before the operation of the classroom was once a domain all but separate from the central office, now data links them.

Piety convincingly argues that education and business, two communities that are often painted as being culturally and substantively separate, are more conceptually linked than we might think.

The next few chapters focus on the use of educational data at different levels of policy making, from the national, to the state, to the district. Piety delights organizational theorists by framing this section of the text by asking the reader to imagine the educational system as having a technical core (where the main work of the organization gets done) and peripheral components (where the managing and tending of the organization happens). Schools do the work of the technical instructional core – here Piety insists that this covers not only classroom instruction, but character building and citizenship developing and socializing that is the product of the entire school experience. The educational data movement has bloated the peripheral components so that they can measure the work of the technical core. But in the best case scenario, it is also providing timely feedback for the technical core with which to improve its practice.

Rarely are we afforded such cogent analysis of a social phenomenon that is happening to us right now. The analysis in the book helps the reader see the landmarks on the short road of the educational data movement, aiding us in understanding how the current data context came to be, and how the ways we think about using data are so dramatically different from just 15 years ago. This kind of reflective narrative history telling is usually reserved for events that are far enough in the past that we have had time and space to process them, or better yet, already seen where the events led and what consequences they had. Piety demonstrates how the educational data movement developed and how it is playing out today with the keen eye of historian even though he is helping us to make sense of our present moment.

Piety demonstrates how the educational data movement developed and how it is playing out today with the keen eye of historian even though he is helping us to make sense of our present moment.

On a more critical note, this book makes no appeals to people who would like to see less data collection and fewer assessments in our schools. There are a large number of stakeholders in the education world who would curtail data collection and standardized testing, if given the chance. They are parents, teachers and educational activists and they believe that children are over tested and that education should be locally controlled and not standardized. None of Piety's arguments respond to any of the anxieties of skeptics of educational data. This is a mistake, because the ideas in the book could help bridge the divide between those communities.

On a more critical note, this book makes no appeals to people who would like to see less data collection and fewer assessments in our schools.

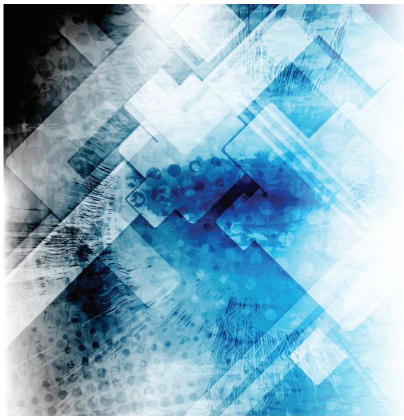
There are some new ideas here that would be applied to higher education assessment. For example, Piety encourages policymakers and practitioners to value "information ecologies," that is, rather than making decisions based on a single achievement score data point, to combine performance data and other representations to allow for informed decision making for each unique context (cf. Milliron, Malcolm, & Kil). In a related point, Piety sees room for growth in the areas of collaboration technologies. In stark contrast with transactional technologies – technology that collects data or provides analysis in a one way direction – collaborative technologies create communities of practice, organizational learning and allow for the two way flow of data. In higher education assessment this would mean thinking more creatively about providing usable data analysis to professors and students to inform their decision making about the current or successive semesters.

Higher education assessment professionals have much to learn from the challenges and notable successes of personnel using big data to shape K–12 education programs. While much of the higher education assessment still uses an AYP–like model (comparing a college to itself from year to year) it is likely that we will be taking cues from the K–12 sector and moving to value–added models (measuring what individuals learn over time). Higher education assessment persons should care about big data because we are all a part this enterprise, and because unlike trends in education that raged for a decade and receded, the use of big educational data is here to stay, and is likely to get bigger.

References

Macfadyen, L. P., Dawson, S., Pardo, A., & Gasevic, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy change. *Research & Practice in Assessment, 9*(2), 17-28.

Milliron, M. D., Malcolm, L., & Kil, D. (2014). Insight and action analytics: Three case studies to consider. *Research & Practice in Assessment, 9*(2), 70-89.



Notes in Brief

The new data sciences of education bring substantial legal, political, and ethical questions about the management of information about learners. This piece provides a synoptic view of recent scholarly discussion in this domain and calls for a proactive approach to the ethics of learning research.

AUTHOR
Mitchell L. Stevens, Ph.D.
Stanford University

An Ethically Ambitious Higher Education Data Science

The work assembled in this issue leaves little doubt that postsecondary assessment is in a sea change. Digitally mediated instruction provides data whose fidelity to processes of learning are superior to any available to this field in the history of quantitative inquiry. The papers and reviews collected here provide a tantalizing early sense of the scientific promise of these new empirics and a glimpse of their implications for the improvement of higher education.

Yet the opening of a vast new scientific frontier is not the only sea change in postsecondary assessment, or even the most important one. During the same few years that digitally mediated instruction has become a data science, the spiraling cost of attending college in the United States has become a political crisis. During these same few years, the goal of raising stubbornly low rates of college completion has become a major priority for prominent philanthropies. And also during these years, the question of what and how much students actually learn in college has become a major research and policy concern. In sum, the emergence of education data science is simultaneous with an accountability revolution in the postsecondary sector, with many new voices in government and business joining researchers and policy analysts in calls for new means of measuring success in higher education.

Educational measurement is political. It changes the way people make sense of the world and what things count as facts and expertise. It changes relationships between those who produce education, pay for it, and regulate it. It makes educational processes comparable that might long have been regarded as distinct and incommensurate. And it produces information about individuals and groups that can be used by third parties to sell products and distribute fateful opportunities and credentials. This is why the educational data streams now available to scientific inquiry must be considered and managed with thoughtful care.

CORRESPONDENCE

Email
mitchell.stevens@stanford.edu

These are the issues that encouraged some 50 educators, scientists, and legal/ethical scholars to convene at the Asilomar Conference Grounds near Monterey, California, in June 2014. Their task was to specify the ethical challenges and obligations that accompany research on higher education in the era of big data. The convening was modeled after a 1975 event at the same site, during which 140 biologists, lawyers and physicians met to write voluntary guidelines for ensuring the safety of recombinant DNA technology. Another precedent was a 1978 meeting at the Belmont Conference Center in Elkridge, Maryland, which produced a document informing ethical considerations of research with human subjects.

The outcome was a heroically brief document affirming the importance of pursuing education data science for the improvement of higher education in an open, urgent, and ethically considered way. The Asilomar participants concurred that the political implications of measurement in higher education should not inhibit its pursuit, since the prospect of improving higher education with new science was too important a goal to inhibit inquiry.

The Asilomar Convention for Learning Research in Higher Education includes two basic tenets:

> **Advance the science of learning for the improvement of higher education.** The science of learning can improve higher education and should proceed through open, participatory, and transparent processes of data collection and analysis that provide empirical evidence for knowledge claims.
>
> **Share.** Maximizing the benefits of learning research requires the sharing of data, discovery, and technology among a community of researchers and educational organizations committed, and accountable to, principles of ethical inquiry held in common.

The Asilomar Convention for Learning Research in Higher Education… specifies six principles to inform decisions about data use and knowledge sharing in the field: Respect for the rights and dignity of learners; beneficence; justice; openness; the humanity of learning; and continuous consideration of the ethical dimensions of learning research.

The Convention additionally specifies six principles to inform decisions about data use and knowledge sharing in the field: *Respect for the rights and dignity of learners; beneficence; justice; openness; the humanity of learning; and continuous consideration* of the ethical dimensions of learning research. The entire document is available at asilomar–highered.info. By way of informing the discussion represented in this issue of *Research & Practice in Assessment*, I add a brief word here about the final principle.

Anyone who pursues education data science quickly learns that there is considerable uncertainty about just how inherited norms and routines for ethical oversight should be applied to data from digitally mediated instruction. IRB protocols that require active consent (rather than a continuous flow of data collection) and prior specification of research questions (rather than iterative inquiry), university proprietary rules that presume data have single owners or trustees (rather than multiple ones), and legal rules applying specifically to students (rather than learners) are but a few features of standard regulatory architecture that fit only awkwardly, if at all, to research with data from digitally mediated instruction. What to do?

Educational measurement is political. It changes the way people make sense of the world and what things count as facts and expertise.

One option would be wait until our IRB officers, attorneys, government and foundation officials, and politicians figure out how to rewrite the inherited rules. In light of the inherent complexity of this problem it is unclear just how long that wait might be. A second option is to move forward with research with an explicit commitment to what the Asilomar Convention calls *continuous consideration*. "In a rapidly evolving field there can be no last word on ethical practice" it reads. "Ethically responsible learner research requires ongoing and broadly inclusive discussion of best practices and comparable standards among researchers, learners, and educational institutions.[1]"

I believe that the second option is by far the ethically more ambitious one. It recognizes the complexity of the current historical moment while keeping sight of the extraordinary opportunity for new science to improve the quality of instruction and learning in college. It recognizes that ongoing peer review is an essential component of responsible scientific conduct. And it enables us to inform the ongoing development of ethical tradition with the wisdom and caution that comes only with practice.

Moving forward quickly and ambitiously with higher education data science will not be uncontroversial. As this mode of inquiry gains intellectual space and analytic sophistication, it will almost surely direct attention away from currently preponderant modes of measuring value in the sector: persistence and completion rates, accreditation review protocols, rating and ranking schemes, and the myriad social sciences of higher education that have been built with student–level survey and census data. Each of these measurement regimes has partisans and profiteers who will pay attention to any change in what counts as valid and reliable assessment. Add all of this to the more general ethical questions confronting use and integration of big data generally, and we have research frontier whose obstacles are hardly for the faint of heart.

Thankfully the work itself is thrilling and the possibilities for educational improvement profound. Hang on, keep moving, and steer.

[1] http://asilomar-highered.info/