Chapter 5

Mining Big Data in Education: Affordances and Challenges

CHRISTIAN FISCHER ID
University of Tübingen

ZACHARY A. PARDOS ID
University of California, Berkeley

RYAN SHAUN BAKER
University of Pennsylvania

JOSEPH JAY WILLIAMS
University of Toronto

PADHRAIC SMYTH
RENZHE YU
University of California, Irvine

STEFAN SLATER
University of Pennsylvania

RACHEL BAKER
MARK WARSCHAUER
University of California, Irvine

The emergence of big data in educational contexts has led to new data-driven approaches to support informed decision making and efforts to improve educational effectiveness. Digital traces of student behavior promise more scalable and finer-grained understanding and support of learning processes, which were previously too costly to obtain with traditional data sources and methodologies. This synthetic review describes the affordances and applications of microlevel (e.g., clickstream data), mesolevel (e.g., text data), and macrolevel (e.g., institutional data) big data. For instance, clickstream data are often used to operationalize and understand knowledge, cognitive strategies, and behavioral

processes in order to personalize and enhance instruction and learning. Corpora of student writing are often analyzed with natural language processing techniques to relate linguistic features to cognitive, social, behavioral, and affective processes. Institutional data are often used to improve student and administrational decision making through course guidance systems and early-warning systems. Furthermore, this chapter outlines current challenges of accessing, analyzing, and using big data. Such challenges include balancing data privacy and protection with data sharing and research, training researchers in educational data science methodologies, and navigating the tensions between explanation and prediction. We argue that addressing these challenges is worthwhile given the potential benefits of mining big data in education.

In recent decades, the increased availability of big data has led to new frontiers in how we monitor, understand, and evaluate processes in educational contexts and has informed decision making and efforts to improve educational effectiveness. Although no single unified definition exists, big data are generally characterized by high volume, velocity, and variety in the digital era (Laney, 2001; Ward & Barker, 2013). Compared with earlier generations of data collected through considerable human effort, the prevalent use of digital tools in everyday life generates an unprecedented amount of data (volume) at an increasing speed (velocity) and from different modalities and time scales (variety; Laney, 2001; Ward & Barker, 2013). Thus, these data require considerable computing resources and alternative analytical methodologies to process and interpret. The National Academy of Education (2017) states that "in the educational context, big data typically take the form of administrative data and learning process data, with each offering their own promise for educational research" (p. 4).

The emergence of big data in education is attributed to at least two major trends in the digital era. First, the recording and storing of institutional data in traditional settings have become increasingly digitized, resulting in vast amounts of standardized student information. Specifically, student information systems (SIS) have been widely adopted to store and organize student profile information (e.g., demographics, academic background) and academic records (e.g., course enrollment and final grades) in schools. These data traditionally encompass decades of students at an institution, with an institution's SIS making it possible to manage and analyze those data at scale. Second, learning behaviors that were challenging to record in face-to-face classrooms can now be partially captured by learning management systems (LMS). In most cases, LMS are used by instructors to distribute instructional materials, manage student assignments, and communicate with students. From clicks on course modules to revisions of an essay submission, these time-stamped logs easily amount to thousands of data points for an individual student. Beyond SIS and LMS, the variety of innovations in digital learning environments enrich new pedagogical possibilities and, in the meantime, collect students' digital footprints. This diversity leads to heterogeneous and multimodal data in large volumes.

A broad range of data mining techniques can be utilized for big data in education, which Baker and Siemens (2014) broadly categorize into prediction methods, including inferential methods that model knowledge as it changes; structure discovery algorithms, with emphasis on discovering the structures of content and skills in an educational domain and the structures of social networks of learners; relationship mining, including sequential pattern mining and correlation mining; visualization; and discovery with models, including using models in subsequent analyses.

With their volume, velocity, and variety, all these "big data" represent a high-value perspective on learner behavior for multiple fields of education research. Questions that were either costly or even impossible to answer before these data sources were available can now be potentially addressed. Digital traces of student actions promise a more scalable and finer-grained understanding of learning processes. By combining behavioral data with surveys or psychological scales, researchers can map action sequences to cognitive traits and test whether observed behavioral traces align with theoretical assumptions and refine theories at a granular level. This rich information has the potential to help understand the mechanisms of specific policy effects and to address policy-relevant issues. For example, connecting administrative and learning process data can unveil nuances about educational inequities and inform actions in faster feedback cycles. The goal of finding effective instructional approaches comparable with one-to-one tutoring has been sought after for decades, and the magnitude of learning process data makes it possible to personalize learning experiences in new ways.

Framework for the Review

This review describes the affordances of big data use in education at three broad levels relevant to educational contexts: the *microlevel* (e.g., clickstream data), *mesolevel* (e.g., text data), and *macrolevel* (e.g., institutional data).

Microlevel big data are fine-grained interaction data with seconds between actions that can capture individual data from potentially millions of learners. Most microlevel data are collected automatically during interactions between learners and their respective learning environments, which include intelligent tutoring systems, massive online open courses (MOOCs), simulations, and games.

Mesolevel big data include computerized student writing artifacts systematically collected during writing activities in a variety of learning environments ranging from course assignments to online discussion forum participation, intelligent tutoring systems, and social media interactions. Notably, mesolevel data affords opportunities to naturally capture raw data on learners' progressions in cognitive and social abilities, as well as affective states.

Macrolevel big data comprise data collected at the institutional level. Examples of macrolevel data include student demographic and admission data, campus services data, schedules of classes and course enrollment data, and college major requirement and degree completion data. While macrolevel data are generally collected over multiyear time spans, they are infrequently updated, often only once or twice per term (e.g., course schedule information, grade records).

Notably, these micro-/meso-/macrolevel categorizations should not be viewed as strictly distinct levels as there can be considerable overlap within each data source. For example, keystroke logs in intelligent tutoring systems represent microlevel data that could provide insights on writing behavior (e.g., burst writing, editing processes). In turn, the content and linguistic features of written texts represent mesolevel data that could be analyzed with natural language processing (NLP) approaches. Similarly, social media interactions often entail microlevel time stamps (and sometimes location information), in addition to the mesolevel contents of each posting. Also, social media data frequently allow researchers to analyze the mesolevel relational positioning between users. Another example is college application materials. Essays are frequently a standard component of university application processes, which provide both mesolevel text data and macrolevel institutional data.

Literature Search

Given the fast-growing nature of relevant research, our synthetic review is primarily based on the literature of the past 5 years (2014–2018), while building on several review and synthesis papers (e.g., Baker & Yacef, 2009; Baker & Siemens, 2014; Pardos, 2017). More specifically, the research communities that examine big data in education increasingly focus on providing policy-relevant insights into education and learning in a variety of learning contexts. Thus, we mostly draw on refereed conference proceedings and peer-reviewed journals from these communities, including the International Conference on Learning Analytics and Knowledge, the International Conference on Educational Data Mining, the International Conference on Artificial Intelligence in Education, the ACM (Association for Computing Machinery) Conference on Learning at Scale, the *International Journal of Artificial Intelligence in Education*, the *Journal of Educational Data Mining, IEEE Transactions on Learning Technologies*, and the *Journal of Learning Analytics*. However, seminal papers from other outlets that are not primarily outlets for big data research (and thus not part of the above list) were also considered based on the authors' expertise in their respective areas.

Papers included for consideration had to be original empirical studies that analyzed real-world data. Thus, papers that described simulation studies, replication studies, and meta-analytic studies were not included in this synthetic review. We did not consider papers that solely report on methodological improvements or conceptual papers. Also, the research needed to be situated in a formal or informal educational context. For instance, research studies that focused on students, teachers, classrooms, learning platforms, schools, or universities were eligible for inclusion in this synthetic review. Regarding analytical strategies, studies that were included needed to have used data mining techniques, rather than just qualitative methods or descriptive statistical analyses. Data needed to be digitally recorded and/or archived at scale. In most cases, this excluded traditionally summative educational data (e.g., surveys, test performance) and new digitized data that were currently less feasible to collect at scale (e.g., data from audio, visual, physiological, and neural sensors).

For each paper, we read the abstract and data set description (if provided) to decide whether they fit the inclusion criteria of this review. Then, the studies were examined to verify that they did not meet the exclusion criteria. The remaining studies were categorized as micro-, meso-, and macrolevel studies. Notably, a study could be assigned more than one category. In total, we identified 370 papers eligible for the section on microlevel big data, 175 for mesolevel big data, and 57 for macrolevel big data, as well as about 200 short papers. Papers included in the list of potentially eligible studies were carefully reviewed by experts on the author team in their respective area of expertise to identify and synthesize larger conceptual themes.

MICROLEVEL BIG DATA

Microlevel big data in education consist of data that can occur at the granularity of seconds between actions. Although multimodal data are increasingly commonly used in learning analytics (Ochoa & Worsley, 2016), the majority of microlevel data used in education consist of data produced by exchanges between learners and data collection platforms in MOOCs, intelligent tutoring systems, simulations, and serious games. This type of data includes information about both the learner's actions and the context in which those actions occur. Often, this type of data is not large in terms of numbers of students—in many cases only hundreds of students are considered—but the volume of data they produce is often quite large, ranging from tens of thousands to millions of data points. In some cases, models are developed for and applied to hundreds of thousands of students, bringing the total data size to billions of data points.

The nature and grain size of microlevel clickstream data make such data well suited to situations where direct intervention might be useful, such as providing students with scaffolding or feedback based on their cognitive or affective states or moving students to a new topic on a knowledge component when they are ready. The scale of clickstream data also facilitates their use across large numbers of contexts and situations, such as studying the development of student learning and engagement over the scale of months or differentiating between student groups who are too rare to show up in small samples.

Microlevel data are often used to detect cognitive strategies, affective states, or self-regulated learning (SRL) behaviors, and they are sometimes validated based on real-time observations of student actions (Botelho et al., 2017; DeFalco et al., 2018; Pardos et al., 2014) or retrospective hand coding of data subsets (Gobert et al., 2012). Then, these detectors are used to study the construct of interest (Pardos et al., 2014; Sao Pedro et al., 2014; Tóth et al., 2014) and drive automated intervention (Aleven et al., 2016; DeFalco et al., 2018; Moussavi et al., 2016). This two-step process necessitates the identification of constructs of interest, either through quantitative coding or by obtaining labels in another fashion (e.g., self-report), and the construction of a machine-learned model that can accurately identify the presence or absence of the construct.

In this section, we review research that used microlevel data to operationalize and understand (a) knowledge components, (b) metacognition and self-regulation, and (c) affective states, as well as to evaluate (d) student knowledge. We also consider how microlevel data mining can identify (e) actionable knowledge to enhance instruction and learning and (f) how to personalize digital educational resources.

Identifying Knowledge Components

There has been considerable prior work on using microlevel data to make inferences about how student performance relates to complex cognitive skills within learning activities. Complex cognition has historically been difficult to infer at scale, but new data mining methods made it possible to model and track it over time. Hundreds of students typically generate vast numbers of interactions, ranging from magnitudes of ten thousand to millions of interactions. Automated detectors that identify students' behavioral patterns have been developed and applied to data sets to identify the degree to which students transferred their knowledge of scientific inquiry between domains and to improve outcomes, driving automated scaffolding aimed at improving students' ability with these skills (Moussavi et al., 2016; Sao Pedro et al., 2014). This work was followed by considerable interest in studying problem-solving strategies. For instance, Tóth et al. (2014) studied problem solving within the MicroDYN learning environment and clustered how student strategies developed and shifted over time. Similarly, Bauer et al. (2017) examined problem-solving approaches in the scientific discovery game *Foldit*, which tasks users with identifying protein structures, a biology research task that is difficult to do in a fully automated fashion. By using visualization to understand the clickstream data produced within the game, the authors identified several common problem-solving strategies and associated these strategies with players' performances. Bauer and colleagues noted that understanding these approaches could be used to provide scaffolding that could improve the quality of players' solutions.

Identifying Metacognitive and SRL Skills

Within the educational data mining community, many researchers have also studied metacognition and SRL. These constructs often examine the learner's ability to self-regulate learning processes (Roll & Winne, 2015), behaviors that are especially relevant in less structured systems such as LMS and MOOCs. Samples ranged from ten to tens of thousands of students and included up to 100 million interactions. Educational data mining approaches to examining SRL often involve modeling the processes and actions that students undertake within learning environments to identify possible scaffolds to encourage learning, which system developers and designers may use to improve user interfaces and experiences (Aleven et al., 2016; Roll & Winne, 2015).

Microlevel clickstream data are uniquely positioned to provide detailed information on students' temporal and sequential patterns of behaviors based on specific

actions students undertake and the system design components students utilize. For instance, Park et al. (2017) explored the development and validation of an effort regulation measure using clickstream data on students' previewing and reviewing of course materials. Students who increased their efforts to review course materials were more likely to pass the course, whereas students who decreased their efforts were less likely to pass the course. Similarly, Park et al. (2018) developed and validated a time management measure that identifies student procrastination and regularity of procrastination based on student clickstream data in online courses with periodic deadlines. Students who received As had significantly higher time management skills (i.e., regular nonprocrastinators) than B grade students (i.e., irregular procrastinators/ irregular nonprocrastinators), who had significantly higher time management skills than C/D/F grade students (i.e., regular procrastinators).

There has also been considerable research into SRL within the Betty's Brain teachable agent and learning management platform for middle school science (Biswas et al., 2016; Segedy et al., 2015). In Betty's Brain, students are tasked with teaching a computer agent (Betty) by producing causal maps and models describing science phenomena. Students' ability to teach Betty is evaluated by a second computer agent, Mr. Davis, who gives Betty quizzes and grades her performance based on how well the student instructed Betty. The Betty's Brain platform provides SRL support to students through both computer agents. For instance, Segedy et al. (2015) clustered SRL behaviors and investigated their associations with student learning in key domain-specific concepts.

Many studies investigated metacognitive and SRL skills in Cognitive Tutors, an intelligent tutoring system for mathematics. A prominent line of SRL research targets help-seeking skills (Aleven et al., 2016). Researchers used microlevel data to develop models of instructional hand-offs (Fancsali et al., 2018), which use student help-seeking behavior and SRL practices to understand how students transition between using different learning resources. For example, Ogan et al. (2015) investigated how help-seeking strategies correlate with learning, using the same learning system and content in different translations. Lu and Hsiao (2016) studied how student behavior during programming correlates to their help seeking within discussion forums and determined that more successful learners read posts in a deeper fashion than less successful learners.

Identifying Affective States

Microlevel data allow us to make inferences about "noncognitive" constructs surrounding engagement, motivation, and affect. The most thoroughly studied constructs are *academic emotions*, also referred to as affective states: frustration, confusion, boredom, and engaged concentration (sometimes called *flow*). Affective states inspired work on developing affect detectors for various learning environments, including intelligent tutoring systems, puzzle games, and first-person simulations (Botelho et al., 2017; DeFalco et al., 2018; Hutt et al., 2019; Pardos et al., 2014;

Sabourin et al., 2011). Detectors are frequently trained on data from hundreds of students with tens of thousands of actions prior to their deployment. Increasingly, this work uses multiple data sources combining quantitative field observations (trained coders observing student behavior during learning and taking systematic notes) and microlevel log data in the development and validation of detectors.

The capacity of educational data mining techniques to identify affective states affords utilization of affective detectors to provide real-time feedback, scaffolding, and interventions to learners. For example, DeFalco et al. (2018) used affective detectors in a military training game to address student frustration as students worked through a combat casualty care skill simulation, TC3Sim, for the U.S. Army. By integrating affective detectors into the game itself, TC3Sim was able to provide feedback messages to students when frustration was identified, leading to improved student learning from pretest to posttest.

Evaluating Student Knowledge

An early application of microlevel clickstream data is the evaluation of student knowledge based on sets of correct and incorrect responses to problems, known as knowledge inference or latent knowledge estimation. Three popular methods are Bayesian knowledge tracing (BKT; Corbett & Anderson, 1995), performance factors analysis (PFA; Pavlik et al., 2009), and deep knowledge tracing (DKT; Khajah et al., 2016). These methodologies use distinct frameworks to infer the degree to which students master given skills. The increasing availability of public data sets such as the Cognitive Tutor and ASSISTments platforms, with data sets often as large as thousands or tens of thousands of students and millions of interactions, has helped this work move forward.

BKT, the oldest of these three approaches, estimates student mastery using a Hidden Markov Model to estimate four parameters for each unique skill contained within the data: the probability that a given student mastered a given skill before the first opportunity to practice that skill; the probability that a student reaches mastery of a skill after the last opportunity to practice but before the next one; the probability that a student who has not mastered a skill will guess on a given opportunity to practice; and the probability that a student who has mastered a skill will answer a given opportunity to practice with an incorrect answer. The parameters of BKT describe qualities of the skill being learned, such as how likely students are to guess at this skill or student prior knowledge. Over the past five years, this framework was expanded to include item difficulty estimates (González-Brenes et al., 2014), answers with partial correctness (Ostrow et al., 2015), and a wider number of possible states for specific knowledge components (Falakmasir et al., 2015). BKT studies support basic research, including on affect detectors, and underpin adaptivity through several learning platforms, such as the Cognitive Tutor (e.g., Liu & Koedinger, 2017).

While BKT uses a Hidden Markov Model to infer student knowledge, PFA (Pavlik et al., 2009) uses logistic regression to estimate three parameters for each

unique skill within the data: the degree to which correct answers are associated with better future performance; the degree to which incorrect answers are associated with better future performance; and the overall ease or difficulty of the skill being estimated. These parameters produce an outcome logit, the probability that a student has mastered a given skill, given the responses up to that point. Compared with BKT, PFA parameters provide less information on the initial knowledge state of learners on a given skill and the predisposition of learners to guess or make careless errors. However, PFA parameters provide insight on the relative difficulty of skills and the relative learning associated with correct and incorrect answers. Extensions of PFA are an active area of research—for instance, to investigate the relative predictive value of recent performance versus older performance (Galyardt & Goldin, 2015), to investigate individual differences in learning rate (Liu & Koedinger, 2015), and to better understand mastery criteria (Käser et al., 2016).

In the past 5 years, DKT has emerged as a popular alternative to BKT and PFA. DKT uses recurrent neural networks to model skill knowledge and mastery, producing a vector of the probability of mastery associated with each opportunity to practice a skill. Compared with the other approaches, DKT is generally more effective at predicting student correctness during learning (Khajah et al., 2016; Yeung & Yeung, 2018), but it has not been used extensively in the real world due to limitations around interpretability and stability of estimates (Yeung & Yeung, 2018).

Using Data for Actionable Knowledge

Big data are also used to understand the effectiveness of administrative decisions and educational interventions. Big data models can predict *when* actions need to be taken for students, such as identifying when students are disengaging from online courses (Le et al., 2018). For instance, Whitehill et al. (2015) analyzed more than 2 million data points generated by more than 200,000 students taking 10 MOOC courses from HarvardX to develop detectors of whether a student would stop course work. These detectors were then used as the basis of interventions that improved student engagement.

In other circumstances, big data have been utilized to discover *what* actions are effective, such as analyzing the larger-scale randomized experiments or randomized controlled trials (Liu et al., 2014; Liu & Koedinger, 2017). Approaches such as reinforcement learning (a subfield of machine learning and artificial intelligence) can create a new paradigm for educational experimentation that attempts to determine which interventions or conditions are effective, and for which students, and to scale those interventions to future students (Liu et al., 2014; Shen & Chi, 2016). Such dynamic experiments estimate the probability that certain conditions are effective, dynamically reweighting randomization so as to present more effective conditions to future students, converging over time to a better instructional policy for each student (Rafferty et al., 2018).

Clustering Student Profiles and Discovering How to Personalize

Actionable knowledge can be gained from assessing which actions are appropriate for different subgroups or profiles of students. Prior research examined hundreds of students in school settings, as well as tens of thousands of students in MOOCs. Examples include identifying how different student groups work through a learning simulation as part of an experimental standardized test (Bergner et al., 2014), modeling how different student groups have different strategies emerge over time in their use of online course resources (Gasevic et al., 2017), and identifying distinct patterns of engagement in MOOCs (Guo & Reinecke, 2014; Kizilcec et al., 2013).

Knowledge of subgroups can inform interventions tailored to different student groups. For instance, recurrent neural networks approaches are used to recommend a timely course page predicted to be relevant to learners given their pattern of engagement (Pardos et al., 2017). Similarly, reinforcement learning can be used to design effective strategies (e.g., problem solving, worked examples) for low- versus high-knowledge learners (Shen & Chi, 2016). These methods have been used to discover how best to sequence practice problems by testing out many different sequences with large numbers of observations from each student (Clement et al., 2015).

Affordances and Challenges of Microlevel Big Data

As this section shows, there are many ways in which microlevel big data have been used in education. Microlevel data are often voluminous, a single student may produce thousands or tens of thousands of data points. It thus becomes possible to analyze phenomena that may take place over a matter of seconds. Affect, for instance, is often detected at a 20-second grain size (Botelho et al., 2017; DeFalco et al., 2018; Pardos et al., 2014), but the resultant detectors can then be used to analyze behavior over the course of an entire year (Pardos et al., 2014; Slater et al., 2016). Analyses at the microlevel lend themselves to models that are relatively easy to apply in interventions. Microlevel big data are, however, not without limitations. Since microlevel big data are easy to collect, many research projects focus solely on them, potentially neglecting important related phenomena that are more coarse-grained. For example, the student knowledge modeling work has focused almost entirely on optimizing immediate prediction, raising possible concerns that these models may be less effective at inferring robust learning that will persist over time (Corbett & Anderson, 1995; Pardos et al., 2014, are notable exceptions). Thus, the ease of collecting microlevel big data does not remove the importance of connecting brief phenomena with longer trends in a learner's development.

MESOLEVEL BIG DATA

Mesolevel big data primarily relate to corpora of writing. The availability of systematically collected computerized student writing artifacts at scale is growing as academic writing moves from paper to digital texts. Whereas one-time national assessments like the ACT/SAT examinations previously constituted a rare opportunity to gather large

writing corpora, submissions of student assignments to LMS made large corpora of writing accessible.

Besides course assignments, textual data can originate from online discussion forums, intelligent tutoring systems, website databases, programming code, and many other sources. Each mesolevel data point is usually collected in time periods that range from minutes to hours. However, an individual may engage in writing activities with varying frequency and regularity. For instance, a student may submit writing assignments every week to LMS over a term to complete a class but may engage in social media interactions with varying intensity over the course of multiple years in the course of a degree program.

Prominent approaches to analyzing text data at scale use NLP tools to automate analytical processes. Linguistic tools can indicate the clusters of lexical, syntactic, or morphological features in student writing; the patterns of collaborative writing in cloud-based corpora; or the quality of student writing normed on corpora of essays previously scored by human graders. For instance, Coh-Metrix (McNamara & Graesser, 2012) reports on linguistics primarily related to text difficulty by measuring components aligned to discourse comprehension including narrativity, syntactic simplicity, word concreteness, referential cohesion, and deep cohesion. Similarly, the Linguistic Inquiry and Word Count tool (Pennebaker et al., 2015) measures psychological constructs including confidence, leadership, authenticity, and emotional tone. Other approaches include social network analysis to generate inferences about relational positionings, and grouping approaches such as *k*-means clustering.

In this section, we review research studies that use mesolevel data to provide insights into (a) cognitive processes (e.g., cognitive functioning, knowledge, and skills), (b) social processes (e.g., discourse and collaboration structures), (c) behavioral processes (e.g., learner engagement and disengagement), and (d) affective processes (e.g., sentiment, motivation).

Supporting and Evaluating Cognitive Functioning

Studies related to cognitive processes have focused on supporting and evaluating learners' cognitive functioning, knowledge, and skills, as well as providing instructors with support (e.g., automated student feedback, automated assignment grading). In recent years, the ability to automate evaluations of student learning expanded from multichoice formats to student writing samples. These studies typically utilize writing samples of hundreds or thousands of students as well as reading comprehension data sets with hundreds of thousands of interactions. Numerous studies demonstrate that evaluation of student writing can be automated to substantially reduce human effort in grading essays in a range of subjects (e.g., Allen et al., 2018; Allen & McNamara, 2015; Head et al., 2017; Lan et al., 2015). For instance, Lan et al. (2015) examined how to automatically grade open-response questions in mathematics. In this work, mathematical solutions for four open-response problems were converted into numerical features, which were then clustered into incorrect, partially correct, and correct solutions. Based on instructor grade assignments for each cluster, the

student solutions were then automatically graded. Studies found students' overall linguistic abilities to be associated with student performance in mathematics and other disciplines (e.g., Crossley et al., 2018; Wang, Yang, et al., 2015). For instance, Crossley et al. (2018) examined the associations between students' mathematical self-concept, interest in mathematics, written interactions with the learning platform, and performance indicators in a blended-learning mathematics program. In particular, Crossley and colleagues found that NLP-derived features were associated with students' mathematical identity (self-concept, interest, value) and mathematics ability. These findings encourage the design of early-warning systems that flag students who are at greater risk of underperforming to instructors. In large lecture courses, these systems may be able to help instructors better identify students who need additional support.

In addition to evaluations of student work, researchers have developed support systems that automate feedback to learners and provided hints to support learning in a variety of domains. For instance, Price et al. (2016) developed a Contextual Tree Decomposition algorithm to provide students working on programming assignments in an intelligent tutoring system with hints on their next steps. These automatically generated hints effectively guide students toward correct solutions of the programming tasks.

Oher research examined how to support instructors with developing assessments by automating the process to evaluate and generate questions. For instance, Wang et al. (2018) used recurrent neural networks models to automatically generate open-response questions from textbooks based on the Stanford Question Answering Dataset. Similarly, Harrak et al. (2018) used clustering approaches on medical school lecture questions to provide instructors with suggestions for in-class feedback.

Supporting and Examining Social Processes

Recent studies analyzed dialogue, discussions, and collaboration patterns from online discussion forums, intelligent tutoring systems, and video transcripts to examine social processes. These studies may use thousands of students with up to a few million interactions. For instance, Hecking et al. (2016) examined MOOC discussion forum data and found that social and semantic structures influenced interaction patterns and community formation processes. Similarly, Gelman et al. (2016) analyzed user interactions on Scratch, an informal learning environment for block-based programming language. Much like in physical spaces, interest-driven subcommunities emerged over time. Besides fully online learning environments, blended-learning formats also provide opportunities for students to engage in collaborative learning. For example, Scheihing et al. (2018) studied a microblogging platform to identify differences in student interaction patterns. In classroom settings, transcript data from video recordings can be used to automate classifications of classroom discourse structures. For instance, Cook et al. (2018) examined classroom recording transcripts, utilizing speech recognition and NLP to detect a characteristic of effective teaching,

the proportion of authentic questions asked in a class session. This finding is mirrored in research that examines and classifies dialogue sequences in intelligent tutoring systems (Dzikovska et al., 2014).

Detecting Behavioral Engagement

Studies related to behavioral engagement analyzed student course engagement and resource-seeking behavior, often utilizing hundreds of thousands of interactions from up to tens of thousands of students. For example, Epp et al. (2017) examined communication behavior in online discussions, with a particular emphasis on student pronoun use. They found that students in instructor-facilitated courses demonstrated higher levels of interaction and used more personal pronouns, whereas students in peer-facilitated courses exhibited lower levels of engagement and used fewer personal pronouns. Atapattu and Falkner (2018) used NLP on MOOC lecture videos to find that discourse features of the lecture video content are related to student interactions with the videos (e.g., pausing, seeking). Joksimović et al. (2015) examined course-related participation patterns of MOOC students on Twitter, Facebook, and blogs. They found that the topics discussed were similar across social media platforms and that the most prominent topics emerged relatively early in the course.

To better support resource-seeking behavior, Yang and Meinel (2014) mined textual metadata from lecture video audio tracks to assist users in their video-browsing and search behavior. Similarly, Peralta et al. (2018) developed a recommendation system that uses metadata to support teachers in the exploration of learning resources on an online platform. Also, Slater et al. (2016) evaluated the quality of mathematics problems that were mostly developed by teachers and submitted to an intelligent tutoring system. Notably, Slater et al. examined students engaging in mathematics problems to detect the relationships between semantic features of the problems and student learning or engagement, which could guide teachers in both their mathematics problem selection in classrooms and their development of new mathematics problems.

Examining Affective Constructs

Studies that investigated affective constructs examined learners' self-concept, sentiment, and motivation while engaging in learning opportunities, often examining hundreds or thousands of students. For instance, Crossley et al. (2018) used data from an online tutoring environment by employing NLP tools to identify the relationships of learners' linguistic ability with their mathematics identity (e.g., math value, interest, and self-concept). Similarly, Allen et al. (2016) utilized NLP to derive the writing characteristics of essays and related them to the affective states of engagement and boredom. In MOOC settings, Wen et al. (2014) utilized discussion forum data in Coursera courses to examine learners' sentiment toward the courses and to identify the relationships between sentiment and course dropout.

Investigating learners' motivations for enrollment in MOOCs, Crues et al. (2018) examined the responses to open-ended questions about course expectations during MOOC enrollment processes and their relationship with age and gender. Using latent Dirichlet allocation and correspondence analysis, they identified 26 reasons for course enrollment, which were associated with learners' age but not their gender. Similarly, Reich et al. (2015) used structural topic modeling to uncover patterns of semantic meaning in unstructured text in order to understand students' enrollment motivation in an educational policy course.

Affordances and Challenges of Mesolevel Big Data

As outlined in this section, mesolevel big data provide several affordances to researchers. Text data can provide insight into students' understanding, their views on various topics, and even their emotional affect. Such data can also give information on relationships and networks within an online community. Studies that use textual analysis may help instructors design courses and activities to improve student engagement and to facilitate peer-to-peer learning (e.g., Atapattu & Falkner, 2018; Gelman et al., 2016; Slater et al., 2016). However, the applicability of various tools (e.g., Coh-Metrix and Linguistic Inquiry and Word Count) has not been tested extensively in all educational settings (Fesler et al., 2019). Researchers cannot ignore contextual factors such as the stimuli to which students are responding. If researchers do not pay attention to unique contextual factors, techniques for analyzing mesolevel big data might result in inaccurate inferences. Such errors are particularly dangerous when tied to important outcomes such as student grades (e.g., Lan et al., 2015).

MACROLEVEL BIG DATA

Macrolevel big data are collected over multiyear time spans, with low rates of collection relative to the other levels. For instance, university-wide institutional data include student demographic and admission data, course enrollment and grade records, course schedule and course descriptions, degree and major requirement information, and campus living data. These data are infrequently updated, at most every few weeks and often only once or twice per term. For instance, student demographic information is usually collected only once and only updated per student request. Nonetheless, such data can afford administrators opportunities to engage in data-driven decision making to improve administrative decision making, enhance student experiences, and improve college or K–12 success.

In this section, we focus on three common application areas of macrolevel data that have emerged in the literature: (a) early-warning systems, also known as early-alert systems; (b) course guidance and information systems; and (c) administration-facing analytics.

Early-Warning Systems

Traditionally, signs that students may be at risk of dropping a course or dropping out of a program are first responded to when students reach out to an instructor or

adviser. The affordance of data-driven early-warning systems is that preemptive support is possible given the availability and utilization of decades of institutional big data, often consisting of tens of thousands of students combined with predictive modeling. Studies assessed real-world deployments of early-warning systems; however, a challenge remains of selecting the appropriate institutional response and types of information to convey to students in order to effectively increase their chances of success (Chaturapruek et al., 2018; Jayaprakash et al., 2014). Notably, a financial evaluation of deployed early-warning systems concluded that setting up early-warning systems and deploying their interventions was cost-effective (Harrison et al., 2016).

Early applications of institutional early-warning systems predicted and responded to course-level failure. Marist College piloted a system that predicted students' likelihood of failing a course based on LMS session data, academic standing, demographics, and standardized test scores (Jayaprakash et al., 2014). Candidate predictive models were trained to predict course failure. The most accurate model was used in a real-time controlled study to trigger an intervention for any students who were predicted to fail a course. For students in the experimental condition, the system dispatched an email alerting them that they were at risk of failing the course and describing resources they could seek to receive support (Harrison et al., 2016; Jayaprakash et al., 2014). The intent of the intervention was to increase the flagged students' chances of success in the course; however, the results were mixed. A statistically significant increase in average course grade of 2 to 5 percentage points was observed in the experimental condition over the control. However, about 7% to 11% more students in the experimental condition withdrew from the course compared with students in the control condition (Jayaprakash et al., 2014).

Course Guidance and Information Systems

Course information and guidance systems have emerged as a complement to early-warning systems. Instead of responding to early signs of trouble in a class, they instead aim to help students select their courses. An example of a deployed system is AskOski at University of California, Berkeley, which uses historic enrollments and machine learning to suggest courses across campus that may be relevant to students' interests and links them to the campus degree audit system to give personalized recommendations of courses that would satisfy students' unmet graduation requirement (Pardos, Fan, et al., 2019). Another deployed system, Stanford's CARTA system, surfaces historic course grade distributions, course evaluations, and common courses taken before and after a course (Chaturapruek et al., 2018). As with the early-warning intervention at Marist, unintended results were observed in CARTA's surfacing of course grade distributions, leading to one-quarter reduction in grade point average (GPA) for students encouraged to use the system. These findings underscore the importance of understanding how different types of information affect student choices, agency, and success.

Off-line experiments applying machine learning to predict student course grades have been increasingly commonplace in the literature (O'Connell et al., 2018; Ren et al., 2017; Sweeney et al., 2016). As data sources and techniques for achieving high accuracy in this prediction task become established, the methodological question shifts toward using models to support students in achieving their desired performance. Nascent work (Jiang et al., 2019) has investigated if recommendations for preparation courses outside of the standard prerequisites can be data mined from historical course enrollment and performance data. Furthermore, degree-level and institution dropout, particularly within the first semester, has been frequently studied (Aguiar et al., 2014; Chen et al., 2018; Gray et al., 2016; Zhang & Rangwala, 2018). For example, Gray et al. (2016) predicted which students are likely to earn a failing-level GPA in the first semester based on course selection, age, and prior academic performance in secondary school. On-time versus over-time graduation expectations have also been modeled. Hutt et al. (2018) predicted college-level outcomes from macrolevel data even before a student arrives on campus. Using a national data set, Hutt and colleagues investigated the use of binary classification models to predict whether students would graduate within 4 years, using 166 features as predictor variables, including student demographics, standardized test scores, academic achievement, and institution-level graduation rates.

Administration-Facing Data Analytics

Méndez et al. (2014) argue that "simple techniques applied to readily-available historical academic data" (p. 148) can provide valuable inside perspectives of educational institutions' programs. Institutional data sets typically contain decades of data from hundreds of thousands of students accumulating millions of course enrollments. Relatively straightforward data visualization, exploration, and -modeling techniques can be quite useful, and more advanced methods are not necessary to extract useful information, although such techniques are less popular in the literature, which often emphasizes the development and application of more complex methodologies. For instance, Méndez and colleagues extracted insights from course outcome data in a computer science program by utilizing the included estimation of course dependence via pairwise linear correlation of grades for the same student across pairs of courses, inference of curriculum coherence via factor analysis of student grades across multiple courses, and identification of dropout paths via sequence mining of the course paths of students who dropped out. This combination of techniques provided insights that were obvious retrospectively but hidden otherwise. For example, many dropouts occurred early in student trajectories due to failing courses in basic science (rather than computer science), suggesting that focusing tutorial resources on these science courses might help increase retention rates. Work has also extended from identifying relationships between courses within an institution to identifying such relationships across institutions. Pardos, Chau, et al. (2019) used classical and neural networks–based natural language techniques to analyze course catalog descriptions and enrollment records from a 2-year and a 4-year institution to

identify similar courses between them. Their investigation attempted to increase the quantity and quality of course pairs, or articulations, where transfer students would be guaranteed course credit. They found that while the course descriptions provided the most powerful signal of similarity, patterns of enrollment around the course (i.e., who took the course and which other courses they took) were nearly as valuable as the descriptions in identifying similarities across institutions.

Koester et al. (2017) aimed for the "transcript of the future" by using macrolevel data to generate a richer description of a student's academic experience as an alternative to traditional GPA and course grade information. They modeled student–grade pairs as linear combinations of student and course fixed effects and explored estimated student and course effects, identifying various aggregated patterns in enrollment and outcome data. This illustrates that even relatively limited institutional data (records of course outcomes for student–course pairs) can potentially provide a wealth of information about students, courses, and majors. Similarly, Mahzoon et al. (2018) focused on information contained in sequences of student course outcomes to build sequential descriptors of student academic performance across terms from college entrance to graduation, providing a basis for visualizations and automatically generated narratives about student trajectories. This approach derived sequential signatures for each student to predict on-time graduation, concluding that temporal information as a student progresses through college is important in predicting student outcomes.

In addition, course information captured in course syllabi and curricula can be mined for potentially insightful information. For example, Sekiya et al. (2015) analyzed computer science degree curricula across 10 U.S. universities, focusing on online syllabi (available from course webpages) for each computer science course. With topic modeling, Sekiya and colleagues automatically extracted clusters of words in the form of topics or "knowledge areas," where each university's syllabus could be characterized as a distribution over knowledge areas. This approach provides a systematic framework for quantitative comparative analysis and visualization of syllabi across universities, leading to insights about emphases in education across different universities—the use of automated text analysis techniques here is essential given the volume and complexity of the data involved. Davis et al. (2018) analyzed learning design components across 177 MOOCs consisting of more than 78,000 learning components (e.g., assets with which learners interact—videos, problems, html pages, etc.). Sequences of activities were abstracted via "lecture → discussion → assessment" by clustering transition probabilities and sequence mining to generate insights about common sequential learning patterns across multiple courses. While this analysis is relatively new, it has the potential to provide novel insights, for example, by linking thematic aspects of course design with measurements of student activity and performance.

Affordances and Challenges of Macrolevel Big Data

This section highlights the promise of bringing more advanced statistical techniques to bear on extant data sets. Universities routinely collect reams of course-taking and student performance data, but until recently these data were rarely used

for institutional reforms or to improve student decision making. By analyzing these data, and making data and analyses available to students, schools can meaningfully improve outcomes. Importantly, public access to these data may also improve equity. Whereas course-taking information was historically available only through social networks, such as fraternities and sororities, more open access may have a democratizing effect by giving all students equal access.

However, benefits of these data sources may be limited in several ways. First, schools' contexts are unique, and applying the same analysis across schools may yield unreliable findings. For example, curricular requirements across majors or schools can affect student course taking, and knowledge of these requirements can affect inferences from analyses. Second, if students have goals not captured by institutional data, such as employment outcomes, the available data may provide limited guidance. Joining multiple sources of data, such as employment records or students' social activities on and off campus, could improve researchers' ability to make inferences but may also raise concerns about student privacy. Finally, as with all types of big data, it is uncertain how students may use the information from these analyses to change their behavior. As Chaturapruek et al. (2018) found, informational interventions may have unintended consequences on student behavior and student outcomes.

CHALLENGES

Though data mining offers numerous potential benefits for education research, there are also many challenges to be overcome to achieve those benefits. We summarize them below in three main areas: accessing, analyzing, and using big data.

Accessing Big Data

Educational data exist in a wide array of formats across an even wider variety of platforms. In almost all cases, these platforms were developed for other purposes, such as instruction or educational administration, rather than for research. Many commercial platform providers, such as educational software companies, have no interest in making their data available publicly. Other companies make their data available in a limited way but have not invested resources to facilitate access to data for research. Only a small number of platforms, such as Cognitive Tutor and ASSISTments, have made high-quality data broadly available.

By contrast, Google makes available the API (Application Programming Interface) of its widely used Google Docs program so that third-party companies can create extensions and other products that use or integrate with the software. It also allows users to view the history of their writing process in individual documents they have written or collaborated on down to 4-second increments; these documents can also be shared with others, who can also view those histories. The combination of open API and document history should, in theory, allow users to analyze metadata from large sets of writing data, for example, all documents written by students and teachers

in a school district under a Google Docs site domain. In principle, though, writing the software to extract and analyze the data is a hugely complicated task. Some university and commercial groups have taken small steps in this direction, including the Hana Ohana research lab at University of California, Irvine, which has developed tools for analyzing collaboration history on individual Google Docs (Wang, Olson, et al., 2015), and the private company Hapara (2019), which mines school district data for patterns related to time and amount of student writing, but these are very partial solutions to what largely remains an out-of-reach treasure of student writing data. In addition, even platforms that make their data available may require programming skills to extract the data. Though many education researchers are familiar with statistical software such as R or Stata, far fewer know programming languages superior for data extraction, such as Python.

Finally, and most important, the availability of data is complicated by privacy issues. Parents, educators, and others are rightly concerned about companies' ability to mine large amounts of sensitive student data and act in ways that are not necessarily focused on bettering individual students' futures. Fears have been raised that student data that are inappropriately shared or sold could be used to stereotype or profile children, contribute to tailored marketing campaigns, or lead to identity theft (Strauss, 2019). Data privacy issues are exacerbated in K–12 settings, where students are children and participation in educational activities is mandatory.

Though the risks of sharing student data generate the most publicity, there are also risks to *not* sharing student data. Colorado has the strictest student data–sharing policies in the United States, according to the Parent Coalition for Student Privacy (2019). Yet data sharing is so strict that, according to the Right to Know (2019) coalition (see also Meltzer, 2019; Schimke, 2019), the public is robbed of the information necessary to evaluate the performance of schools and educational programs in the state and their impact on diverse students.

Finding the right balance between individual privacy and the public interest is very challenging. This is, in part, because the large amount of data available in big data sets makes it very difficult to prevent the "reidentification" of de-identified data, even if all direct identifiers are removed. It is thus impossible to combine maximal privacy with maximal utility. Instead, educational institutions and researchers face a choice between maximizing privacy and limiting the utility of the data set or maximizing utility but leaving the data subject to possible reidentification with sufficient effort (Nelson, 2015).

The challenges of sharing mesolevel data are even greater, since there is an unlimited number of ways in which students can reveal their identity in their writing. Addressing these challenges requires different kinds of strategies for different audiences and purposes. The U.S. Family Education and Privacy Act allows schools and institutions to share data with organizations conducting studies for the purpose of improving instruction. Organizations such as the Inter-University Consortium for Political and Social Research host data sets with a wide range of restrictions. Data sets that favor utility (but sacrifice maximal privacy) can be made available to other

research teams that are governed by institutional review board protocols, while data sets that limit utility but maximize privacy can be shared with the general public. Of course, even groups that are inclined to make data available for research may be hesitant to do so due to the extra steps and expenses required to ensure an appropriate level of de-identification.

Analyzing Big Data

As with accessing big data, analyzing big data also poses challenges regarding researchers' skills. As noted above, few education researchers know key programming languages used for data science, such as Python. Education research graduate programs seldom offer instruction in the data-clustering, -modeling, and prediction techniques used to analyze big data.

Even for researchers with such skills, error rates and noise pose additional challenges. For example, although predictive models can provide systematic improvements in prediction quality on average over base rates, high error rates may indicate the occurrence of significant exogenous factors at play not captured even in large amounts of data. When such predictive results facilitate the decision making of instructors or institutional policymakers, these errors may harm students' short-term learning or long-term success. In addition, large data sets with large numbers of predictor variables may result in models that are quite complex and difficult to interpret and that may not necessarily help stakeholders more than simpler models. This suggests that predicting student outcomes at a macro, "long time scale" level is inherently difficult and relationships between predictors and "downstream outcomes" can be complex, with many different factors affecting student outcomes that may potentially not be measured.

One way to mitigate these challenges is to combine macrolevel data with micro- or mesolevel data. For instance, Aguiar et al. (2014) exemplified how nonmacro data can be useful in predicting student outcomes. The authors investigated different data sources for predicting student dropout of engineering courses at Notre Dame after their first term, treated as a binary classification problem. In terms of institutional (macro) data sources, the authors used predictor variables based on academic performance (i.e., SAT scores, first-term GPA) and demographics (i.e., gender, income group). Microlevel predictor variables included online student engagement during the first college term. The results were strikingly clear: Online engagement variables had significantly more predictive power than academic performance or demographic variables across a variety of classification models. Similarly, Miller et al. (2015) found that predictive models constructed to predict learning outcomes for students taking undergraduate computer science courses could benefit significantly from including online student interaction data. These studies indicated that the addition of predictors based on noninstitutional data (e,g., online engagement data) can provide significant additional predictive power beyond that of institutional data alone.

Using Big Data

Finally, even if we successfully access and analyze big data, additional issues arise related to how such data are used. As education researchers increasingly turn to data mining, they will have to confront the tension between explanation and prediction. Yarkoni and Westfall (2017) discuss this tension in detail in relationship to the field of psychology. They argue that psychology's focus on explaining the causes of behavior has led the field to be populated by research programs that provide intricate theories but have little ability to accurately predict future behaviors. They further suggest that increased focus on prediction using data mining and machine learning techniques can ultimately lead to a greater understanding of behavior.

We also believe that this is true in education research, as seen in the example of Connor's (2019) research on her Assessment2Instruction (A2i) professional support system for reading instruction. Literacy research has been marked by the so-called reading wars between advocates of code-focused (e.g., phonics) versus meaning-focused (e.g., comprehension) instruction. Though a consensus has emerged over time on the critical value of the former, how much it should be supplemented by the latter is a continued debate. Connor's team tackled this issue in a highly creative way, adding a less-talked-about but also important question: Are elementary students best served by individualized (child managed) or whole-class (teacher managed) instruction?

The research team collected vast amounts of data on how much time children spent in (a) code- versus meaning-focused and (b) child- versus teacher-managed reading instruction, as well as (c) children's progression in reading proficiency throughout the year. Data mining techniques were used to develop and refine models indicating what combinations of instruction work best for children at different levels of proficiency and at different points in the school year (Connor, 2019). These models were developed into a software recommender system (A2i) that would assist teachers with grouping students to receive the types of instruction best suited to their needs. Randomized controlled trials were used to compare reading achievement in classrooms using A2i with that in classrooms teaching reading without it, finding strong positive effects for the former. This project thus not only built a valuable predictive tool that can guide teachers and improve literacy outcomes but also added explanatory value as to the differential contributions of code- versus meaning-focused and child- versus teacher-managed instruction.

Finally, in using big data, it is critically important to examine and address potential issues of bias, particularly when algorithms associated with big data lead to predictions and/or policy. For example, much attention has been focused on the potential for racial bias in predictive algorithms used in policing (e.g., Brantingham et al., 2018). The European Union Agency for Fundamental Rights (2018) provides a well-justified set of recommendations for how to minimize bias in big data–derived algorithms. These include ensuring maximum transparency in the development of algorithms, conducting fundamental rights impact assessments to identify potential

biases and abuses in the application of and output from algorithms, checking the quality of data collected and used, and ensuring that the development and operation of the algorithm can be meaningfully explained.

Recommendations

Meeting these challenges will require rethinking both how we develop education researchers and the kinds of research practices our research community favors. Curricula in graduate schools of education overwhelmingly favor research methods that fall within one of two major paradigms: quantitative measurement and hypothesis testing or interpretive qualitative research. Analyzing big data draws on an alternate research paradigm to those used in computational social sciences. Only a handful of doctoral programs in education offer the kinds of research training necessary to develop the educational data sciences of the future, and even fewer offer instruction related to the ethical, moral, and privacy dimensions of working with big data. Partnering with other programs across campus, from computer science, data science, or other fields, is a possibility, but in most universities, there is too little interdisciplinary training across these fields and education. In addition, both faculty and graduate students in computer science and data science are incentivized to focus their research on original contributions to important theoretical challenges and techniques in those fields, rather than on applications of data science in other areas, such as education.

To address this challenge, we need to create broader pipelines of talented data scientists focused on education research. This can be through curricular reform within education graduate programs and/or improved interdisciplinary training across the education and computer/data science fields. Federally funded doctoral and postdoctoral training programs in educational sciences would be one very valuable step in this direction.

Mining big data in education challenges not only how we prepare education researchers, but also what kinds of research practices we engage in. Traditional models of education research privilege the sole author, who gets extra rewards in the hiring, tenure, and promotion process; discourage collaboration between junior and senior scholars because such collaboration taints junior scholars as supposedly lacking independence; and favor hoarding of data, so that investigators reap all the rewards from the data without diminishing their value through sharing. In contrast, research projects that involve data mining typically privilege team science, with junior and senior scholars, and open science, so that large data sets can be combined and reused for new analyses and replication. Of course, there are many reasons to support open science even within the traditional quantitative and qualitative education research paradigms, but the value of adopting open science practices is even more pressing as we transition to conducting more educational data science.

The Sloan Equity and Inclusion in STEM Introductory Courses, launched by the University of Michigan, exemplifies the value of open science for new kinds of education research. Faculty at 10 large research universities connect through parallel and

combined data analyses and continuous exchange of speakers and graduate student researchers to explore and improve instructional practices and outcomes in foundational STEM (science, technology, engineering, and mathematics) courses reaching hundreds of thousands of students. Open sharing of data and team science will be hallmarks of this important research initiative. Perhaps not surprisingly, the project was initiated by a professor of physics and astronomy, a discipline where large-scale open team science is much more common than in education.

CONCLUSION

The availability of big data offers exciting new threads of research and the opportunity to add additional perspective to existing threads in education. All types of big data in education offer affordances and challenges. The sheer amount of microlevel data make big data methods a powerful tool for analyzing learner processes, but that power can lead researchers to ignore broader and potentially more important patterns that cannot be measured at the microlevel. Mesolevel data provide a deep window into cognitive processes by examining individuals' writing, but they are prone to many of the broader challenges of using automated tools for writing measurement (e.g., Raczynski & Cohen, 2018). Macrolevel data can be valuable for taking the broadest look at student persistence and achievement, but the smaller size and coarse measurements of macrolevel data sets may make it difficult to identify the finer-grained mechanisms at play (e.g., Scott-Clayton, 2015).

The limitations of each of these types of big data can be minimized, and the benefits amplified, if future research is triangulated either with the remaining types of big data or with more traditional forms of quantitative or qualitative analysis. Through recording, accessing, analyzing, and utilizing multiple types of data, we can better understand and respond to individual learner behavior as it manifests in the increasingly pervasive digital realm. Furthermore, the ubiquity of big data suggests an increased emphasis on preparing students in educational graduate programs to utilize data science methods as well as a committed push toward open science and research structures that favor collaborative teams, to improve our field's capacity for mining big data for education research. Given the potential benefits of mining big data in education, it is worth our effort to begin addressing these challenges.

ACKNOWLEDGMENTS

ORCID iDs

Christian Fischer [ID] https://orcid.org/0000-0002-8809-2776
Zachary A. Pardos [ID] https://orcid.org/0000-0002-6016-7051

REFERENCES

Aguiar, E., Chawla, N. V., Brockman, J., Ambrose, G. A., & Goodrich, V. (2014). Engagement vs performance: Using electronic portfolios to predict first semester engineering student retention. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 103–112). Association for Computing Machinery. https://doi.org/10.1145/2567574.2567583

Aleven, V., Roll, I., McLaren, B. M., & Koedinger, K. R. (2016). Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *International Journal of Artificial Intelligence in Education*, *26*(1), 205–223. https://doi.org/10.1007/s40593-015-0089-1

Allen, L., Likens, A. D., & McNamara, D. S. (2018). A multi-dimensional analysis of writing flexibility in an automated writing evaluation system. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 380–388). Association for Computing Machinery. https://doi.org/10.1145/3170358.3170404

Allen, L., & McNamara, D. S. (2015). You are your words: Modeling students' vocabulary knowledge with natural language processing tools. In *Proceedings of the 8th International Conference on Educational Data Mining* (pp. 258–265), Madrid, Spain. https://pdfs.semanticscholar.org/3f55/697498e911f1a020ad3b75ba6982e16fe69b.pdf

Allen, L., Mills, C., Jacovina, M. E., Crossley, S., D'Mello, S., & McNamara, D. S. (2016). Investigating boredom and engagement during writing using multiple sources of information: The essay, the writer, and keystrokes. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 114–123). Association for Computing Machinery. https://doi.org/10.1145/2883851.2883939

Atapattu, T., & Falkner, K. (2018). Impact of lecturer's discourse for student video interactions: Video learning analytics case study of MOOCs. *Journal of Learning Analytics*, *5*(3), 182–197. https://doi.org/10.18608/jla.2018.53.12

Baker, R. S., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (2nd ed., pp. 253–274). Cambridge University Press.

Baker, R., & 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. https://doi.org/10.5281/zenodo.3554657

Bauer, A., Flatten, J., & Popović, Z. (2017). Analysis of problem-solving behavior in open-ended scientific-discovery game challenges. In X. Hu, T. Barnes, A. Hershkovitz, & L. Paquette (Eds.), *Proceedings of the 10th International Conference on Educational Data Mining* (pp. 32–39), Wuhan, China. https://pdfs.semanticscholar.org/fa02/2ca5d9b1f5364d5346ce0a6ee1cff0976840.pdf

Bergner, Y., Shu, Z., & von Davier, A. (2014). Visualization and confirmatory clustering of sequence data from a simulation-based assessment task. In *Proceedings of the 7th International Conference on Educational Data Mining* (pp.177–184). International Educational Data Mining Society. https://pdfs.semanticscholar.org/b0da/eb27f2982ced-8ab9fc6646edc33abad04742.pdf

Biswas, G., Segedy, J. R., & Bunchongchit, K. (2016). From design to implementation to practice a learning by teaching system: Betty's Brain. *International Journal of Artificial Intelligence in Education*, *26*(1), 350–364. https://doi.org/10.1007/s40593-015-0057-9

Botelho, A. F., Baker, R. S., & Heffernan, N. T. (2017). Improving sensor-free affect detection using deep learning. In E. André, R. Baker, X. Hu, M. T. Rodrigo, & B. du Boulay (Eds.), *Artificial intelligence in education* (Vol. 10331, pp. 40–51). Springer. https://doi.org/10.1007/978-3-319-61425-0_4

Brantingham, P. J., Valasik, M., & Mohler, G. O. (2018). Does predictive policing lead to biased arrests? Results from a randomized controlled trial. *Statistics and Public Policy*, *5*(1), 1–6. https://doi.org/10.1080/2330443X.2018.1438940

Chaturapruek, S., Dee, T. S., Johari, R., Kizilcec, R. F., & Stevens, M. L. (2018). How a data-driven course planning tool affects college students' GPA: Evidence from two field experiments. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*. Association for Computing Machinery. https://doi.org/10.1145/3231644.3231668

Chen, Y., Johri, A., & Rangwala, H. (2018). Running out of STEM: A comparative study across STEM majors of college students at-risk of dropping out early. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 270–279). Association for Computing Machinery. https://doi.org/10.1145/3170358.3170410

Clement, B., Roy, D., Oudeyer, P.-Y., & Lopes, M. (2015). Multi-armed bandits for intelligent tutoring systems. *Journal of Educational Data Mining*, *7*(2), 20–48. https://doi.org/10.5281/zenodo.3554667

Connor, C. M. (2019). Using technology and assessment to personalize instruction: Preventing reading problems. *Prevention Science*, *20*(1), 89–99. https://doi.org/10.1007/s11121-017-0842-9

Cook, C., Olney, A. M., Kelly, S., & D'Mello, S. (2018). An open vocabulary approach for estimating teacher use of authentic questions in classroom discourse. In *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 116–126), Raleigh, NC. International Educational Data Mining Society. https://pdfs.semanticscholar.org/1a75/8415f51474b9943faea5b506f98cadba1d15.pdf

Corbett, A. T., & Anderson, J. R. (1995). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modelling and User-Adapted Interaction*, *4*(4), 253–278. https://doi.org/10.1007/BF01099821

Crossley, S., Ocumpaugh, J., Labrum, M., Bradfield, F., Dascalu, M., & Baker, R. S. (2018). Modeling math identity and math success through sentiment analysis and linguistic features. In *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 11–20), Raleigh, NC. International Educational Data Mining Society. https://pdfs.semanticscholar.org/6153/3632ce351c2b34486645dec7ea58d7186289.pdf

Crues, R. W., Bosch, N., Anderson, C. J., Perry, M., Bhat, S., & Shaik, N. (2018). Who they are and what they want: Understanding the reasons for MOOC enrollment. In *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 177–186), Raleigh, NC. International Educational Data Mining Society. https://pdfs.semanticscholar.org/a794/99a5e1e87a7bdbdf70df2127d78f5cc0182d.pdf

Davis, D., Seaton, D., Hauff, C., & Houben, G.-J. (2018). Toward large-scale learning design: Categorizing course designs in service of supporting learning outcomes. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3231644.3231663

DeFalco, J. A., Rowe, J. P., Paquette, L., Georgoulas-Sherry, V., Brawner, K., Mott, B. W., Baker, R. S., & Lester, J. C. (2018). Detecting and addressing frustration in a serious game for military training. *International Journal of Artificial Intelligence in Education*, *28*(2), 152–193. https://doi.org/10.1007/s40593-017-0152-1

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

Epp, C. D., Phirangee, K., & Hewitt, J. (2017). Talk with me: Student behaviours and pronoun use as indicators of discourse health across facilitation methods. *Journal of Learning Analytics*, *4*(3), 47–75. https://doi.org/10.18608/jla.2017.43.4

European Union Agency for Fundamental Rights. (2018). *#BigData: Discrimination in data-supported decision making*. https://doi.org/10.1163/2210-7975_HRD-9992-20180020

Falakmasir, M., Yudelson, M., Ritter, S., & Koedinger, K. (2015). Spectral Bayesian knowledge tracing. In *Proceedings of the 8th International Conference on Educational Data Mining*. International Educational Data Mining Society. https://pdfs.semanticscholar.org/0e15/52be96b2d2829c5c62288851b9b139e3e9b6.pdf

Fancsali, S. E., Yudelson, M. V., Berman, S. R., & Ritter, S. (2018). Intelligent instructional hand offs. In *Proceedings of the 11th International Conference on Educational Data Mining* (pp. 198–207), Raleigh, NC. International Educational Data Mining Society. https://pdfs.semanticscholar.org/9bf6/1bd5c6a3a7f65e52a4a8ec8928857b5a1af6.pdf

Fesler, L., Dee, T., Baker, R., & Evans, B. (2019). Text as data methods for education research. *Journal of Research on Educational Effectiveness*, *12*(4), 707–727. https://doi.org/10.1080/19345747.2019.1634168

Galyardt, A., & Goldin, I. (2015). Move your lamp post: Recent data reflects learner knowledge better than older data. *Journal of Educational Data Mining*, *7*(2), 83–108. https://doi.org/10.5281/zenodo.3554671

Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting learning strategies with analytics: Links with self-reported measures and academic performance. *Journal of Learning Analytics*, *4*(2), 113–128. https://doi.org/10.18608/jla.2017.42.10

Gelman, B. U., Beckley, C., Johri, A., Domeniconi, C., & Yang, S. (2016). Online urbanism: Interest-based subcultures as drivers of informal learning in an online community. In *Proceedings of the Third ACM Conference on Learning at Scale* (pp. 21–30). Association for Computing Machinery. https://doi.org/10.1145/2876034.2876052

Gobert, J. D., Sao Pedro, M. A., Baker, R. S. J. D., Toto, E., & Montalvo, O. (2012). Leveraging educational data mining for real-time performance assessment of scientific inquiry skills within microworlds. *Journal of Educational Data Mining*, *4*(1), 111–143. https://doi.org/10.5281/zenodo.3554645

González-Brenes, J., Huang, Y., & Brusilovsky, P. (2014). General features in knowledge tracing to model multiple subskills, temporal item response theory, and expert knowledge. In *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 84–91), University of Pittsburgh. International Educational Data Mining Society. https://pdfs.semanticscholar.org/0002/fab1c9f0904105312031cdc18dce358863a6.pdf

Gray, G., McGuinness, C., Owende, P., & Hofmann, M. (2016). Learning factor models of students at risk of failing in the early stage of tertiary education. *Journal of Learning Analytics*, *3*(2), 330–372. https://doi.org/10.18608/jla.2016.32.20

Guo, P. J., & Reinecke, K. (2014). Demographic differences in how students navigate through MOOCs. In *Proceedings of the First ACM Conference on Learning at Scale Conference* (pp. 21–30). Association for Computing Machinery. https://doi.org/10.1145/2556325.2566247

Hapara. (2019). *Hapara Analytics*. https://hapara.com/analytics/

Harrak, F., Bouchet, F., Luengo, V., & Gillois, P. (2018). PHS profiling students from their questions in a blended learning environment. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 102–110). Association for Computing Machinery. https://doi.org/10.1145/3170358.3170389

Harrison, S., Villano, R., Lynch, G., & Chen, G. (2016). Measuring financial implications of an early alert system. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 241–248). Association for Computing Machinery. https://doi.org/10.1145/2883851.2883923

Head, A., Glassman, E., Soares, G., Suzuki, R., Figueredo, L., D'Antoni, L., & Hartmann, B. (2017). Writing reusable code feedback at scale with mixed-initiative program synthesis.

In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale—L@S '17* (pp. 89–98). https://doi.org/10.1145/3051457.3051467

Hecking, T., Chounta, I.-A., & Hoppe, H. U. (2016). Investigating social and semantic user roles in MOOC discussion forums. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 198–207). Association for Computing Machinery. https://doi.org/10.1145/2883851.2883924

Hutt, S., Gardener, M., Kamentz, D., Duckworth, A. L., & D'Mello, S. K. (2018). Prospectively predicting 4-year college graduation from student applications. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 280–289). Association for Computing Machinery. https://doi.org/10.1145/3170358.3170395

Hutt, S., Grafsgaard, J. F., & D'Mello, S. K. (2019). Time to scale: Generalizable affect detection for tens of thousands of students across an entire school year. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. https://doi.org/10.1145/3290605.3300726

Jayaprakash, S. M., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D. (2014). Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics*, *1*(1), 6–47. https://doi.org/10.18608/jla.2014.11.3

Jiang, W., Pardos, Z. A., & Wei, Q. (2019). Goal-based course recommendation. In *Proceedings of the 9th International Conference on Learning Analytics and Knowledge* (pp. 36–45). Association for Computing Machinery. https://doi.org/10.1145/3303772.3303814

Joksimović, S., Kovanović, V., Jovanović, J., Zouaq, A., Gasevic, D., & Hatala, M. (2015). What do cMOOC participants talk about in social media? A topic analysis of discourse in a cMOOC. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 156–165). Association for Computing Machinery. https://doi.org/10.1145/2723576.2723609

Käser, T., Klingler, S., & Gross, M. (2016). When to stop? Towards universal instructional policies. In *Proceedings of the Sixth International Conference on Learning Analytics and Knowledge* (pp. 289–298). Association for Computing Machinery. http://dx.doi.org/10.1145/2883851.2883961

Khajah, M., Lindsey, R. V., & Mozer, M. C. (2016). How deep is knowledge tracing? In *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 94–101), Raleigh, NC. International Educational Data Mining Society. https://arxiv.org/pdf/1604.02416.pdf

Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (p. 170). Association for Computing Machinery. https://doi.org/10.1145/2460296.2460330

Koester, B. P., Fogel, J., Murdock, W., Grom, G., & McKay, T. A. (2017). Building a transcript of the future. In *Proceedings of the Seventh International Learning Analytics and Knowledge Conference* (pp. 299–308). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027418

Lan, A. S., Vats, D., Waters, A. E., & Baraniuk, R. G. (2015). Mathematical language processing: Automatic grading and feedback for open response mathematical questions. In *Proceedings of the Second ACM Conference on Learning at Scale* (pp. 167–176). Association for Computing Machinery. https://doi.org/10.1145/2724660.2724664

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*(70), 1.

Le, C. V., Pardos, Z. A., Meyer, S. D., & Thorp, R. (2018). Communication at scale in a MOOC using predictive engagement analytics. In M. Mavrikis, K. Porayska-Pomsta, & R. Luckin (Eds.), *Proceedings of the 19th International Conference on Artificial Intelligence in Education* (pp. 239–252). Springer. https://doi.org/10.1007/978-3-319-93843-1_18

Liu, R., & Koedinger, K. R. (2015). Variations in learning rate: Student classification based on systematic residual error patterns across practice opportunities. In *Proceedings of the 8th International Conference on Educational Data Mining*. International Educational Data Mining Society. https://pdfs.semanticscholar.org/0b57/0c5a732552bb50d80eafe1122a3 ef52bad23.pdf

Liu, R., & Koedinger, K. R. (2017). Closing the loop: Automated data-driven cognitive model discoveries lead to improved instruction and learning gains. *Journal of Educational Data Mining*, *9*(1), 25–41. https://doi.org/10.5281/zenodo.3554625

Liu, Y. E., Mandel, T., Brunskill, E., & Popovic, Z. (2014). Trading off scientific knowledge and user learning with multi-armed bandits. In *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 161–168). International Educational Data Mining Society. https://pdfs.semanticscholar.org/029a/3f34047890e2a399094c3755c0c 162aae664.pdf

Lu, Y., & Hsiao, I. H. (2016). Seeking programming-related information from large scaled discussion forums, help or harm? In *Proceedings of the 9th International Conference on Educational Data Mining*. International Educational Data Mining Society. https://pdfs .semanticscholar.org/dc76/33aae5a8a99c2abac1c927617187cfefa7eb.pdf

Mahzoon, M. J., Maher, M. L., Eltayeby, O., Dou, W., & Grace, K. (2018). A sequence data model for analysing temporal patterns of student data. *Journal of Learning Analytics*, *5*(1), 55–74. https://doi.org/10.18608/jla.2018.51.5

McNamara, D. S., & Graesser, A. C. (2012). Coh-Metrix: An automated tool for theoretical and applied natural language processing. In P. M. McCarthy & C. Boonthum-Denecke (Eds.), *Applied natural language processing and content analysis: Identification, investigation and resolution* (pp. 188–205). IGI Global.

Méndez, G., Ochoa, X., & Chiluiza, K. (2014). Techniques for data-driven curriculum analysis. In *Proceedings of the Fourth International Conference on Learning Analytics and Knowledge* (pp. 148–157). Association for Computing Machinery. https://doi.org/ 10.1145/2567574.2567591

Meltzer, E. (2019, May 29). *How are Colorado schools doing? Advocates say the state still holds back too much data*. https://chalkbeat.org/posts/co/2019/05/29/how-are-colorado-schools-doing-advocates-say-the-state-still-holds-back-too-much-data/

Miller, L. D., Soh, L.-K., Samal, A., Kupzyk, K., & Nugent, G. (2015). A comparison of educational statistics and data mining approaches to identify characteristics that impact online learning. *Journal of Educational Data Mining*, *7*(3), 117–150.

Moussavi, R., Gobert, J., & Sao Pedro, M. (2016). The effect of scaffolding on the immediate transfer of students' data interpretation skills within science topics. In *Proceedings of the 12th International Conference of the Learning Sciences* (pp. 1002–1005), Singapore. International Society of the Learning Sciences. https://doi.dx.org/10.22318/icls2016.157

National Academy of Education. (2017). *Big data in education: Balancing the benefits of educational research and student privacy: Workshop summary.*

Nelson, G. S. (2015). Practical implications of sharing data: A primer on data privacy, anonymization, and de-identification. In *Proceedings of the SAS® Global Forum 2015 Conference*. SAS Institute. https://support.sas.com/resources/papers/proceedings15/1884-2015.pdf

Ochoa, X., & Worsley, M. (2016). Augmenting learning analytics with multimodal sensory data. *Journal of Learning Analytics*, *3*(2), 213–219. https://doi.org/10.18608/ jla.2016.32.10

O'Connell, K., Wostl, E., Crosslin, M., Berry, T. L., & Grover, J. P. (2018). Student ability best predicts final grade in a college algebra course. *Journal of Learning Analytics*, *5*(3), 167–181. https://doi.org/10.18608/jla.2018.53.11

Ogan, A., Walker, E., Baker, R., Rodrigo, Ma. M. T., Soriano, J. C., & Castro, M. J. (2015). Towards understanding how to assess help-seeking behavior across cultures. *International*

Journal of Artificial Intelligence in Education, *25*(2), 229–248. https://doi.org/10.1007/s40593-014-0034-8

Ostrow, K., Donnelly, C., Adjei, S., & Heffernan, N. (2015). Improving student modeling through partial credit and problem difficulty. In *Proceedings of the Second ACM Conference on Learning at Scale* (pp. 11–20). Association for Computing Machinery. https://doi.org/10.1145/2724660.2724667

Pardos, Z. A. (2017). Big data in education and the models that love them. *Current Opinion in Behavioral Sciences*, *18*, 107–113. https://doi.org/10.1016/j.cobeha.2017.11.006

Pardos, Z. A, Baker, R. S. J. D., San Pedro, M., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end-of-year learning outcomes. *Journal of Learning Analytics*, *1*(1), 107–128. https://doi.org/10.18608/jla.2014.11.6

Pardos, Z. A., Chau, H., & Zhao, H. (2019). Data-assistive course-to-course articulation using machine translation. In J. C. Mitchell & K. Porayska-Pomsta (Eds.), *Proceedings of the 6th ACM Conference on Learning at Scale*. Association for Computing Machinery. https://doi.org/10.1145/3330430.3333622

Pardos, Z. A., Fan, Z., & Jiang, W. (2019). Connectionist recommendation in the wild: On the utility and suitability of neural networks for personalized course guidance. *User Modeling and User-Adapted Interaction*, *29*(2), 487–525. https://doi.org/10.1007/s11257-019-09218-7

Pardos, Z. A., Tang, S., Davis, D., & Le, C. V. (2017). Enabling real-time adaptivity in MOOCs with a personalized next-step recommendation framework. In C. Thille & J. Reich (Eds.), *Proceedings of the 4th ACM Conference on Learning at Scale* (pp. 23–32). Association for Computing Machinery. https://doi.org/10.1145/3051457.3051471

Parent Coalition for Student Privacy. (2019). *The state student privacy report card: Grading the states on protecting student data privacy*. https://www.studentprivacymatters.org/wp-content/uploads/2019/01/The-2019-State-Student-Privacy-Report-Card.pdf

Park, J., Denaro, K., Rodriguez, F., Smyth, P., & Warschauer, M. (2017). Detecting changes in student behavior from clickstream data. In *Proceedings of the Seventh International Conference on Learning Analytics and Knowledge* (pp. 21–30). Association for Computing Machinery. https://doi.org/10.1145/3027385.3027430

Park, J., Yu, R., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). Understanding student procrastination via mixture models. In *Proceedings of the 11th International Conference on Educational Data Mining*. International Educational Data Mining Society. https://pdfs.semanticscholar.org/67a4/027404642fad7f9baf0c5d76dea14c99f563.pdf

Pavlik, P., Cen, H., & Koedinger, K. R. (2009). Performance factors analysis: A new alternative to knowledge tracing. In *Proceedings of the 2009 Conference on Artificial Intelligence in Education* (pp. 531–538). Association for Computing Machinery. https://dl.acm.org/doi/10.5555/1659450.1659529

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. University of Texas-Austin.

Peralta, M., Alarcon, R., Pichara, K., Mery, T., Cano, F., & Bozo, J. (2018). Understanding learning resources metadata for primary and secondary education. *IEEE Transactions on Learning Technologies*, *11*(4), 456–467. https://doi.org/10.1109/TLT.2017.2766222

Price, T. W., Dong, Y., & Barnes, T. (2016). Generating data-driven hints for open-ended programming. In *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 191–198), Raleigh, NC. International Educational Data Mining Society. https://pdfs.semanticscholar.org/7800/c7981659f2fafbe1a6966ce43f7ca9460f48.pdf

Raczynski, K., & Cohen, A. (2018). Appraising the scoring performance of automated essay scoring systems—Some additional considerations: Which essays? Which human raters? Which scores? *Applied Measurement in Education*, *31*(3), 233–240. https://doi.org/10.1080/08957347.2018.1464449

Rafferty, A. N., Ying, H., & Williams, J. J. (2018). Bandit assignment for educational experiments: Benefits to students versus statistical power. In C. Penstein Rosé, M. Mavrikis, R. Martínez-Maldonado, K. Porayska-Pomsta, H. U. Hoppe, B. McLaren, R. Luckin, & B. du Boulay (Eds.), *Artificial intelligence in education* (Vol. 10948, pp. 286–290). Association for Computing Machinery. https://doi.org/10.1007/978-3-319-93846-2_53

Reich, J., Tingley, D., Leder, J., Roberts, M. E., & Stewart, B. M. (2015). Computer-assisted reading and discovery for student-generated text in massive open online courses. *Journal of Learning Analytics*, *2*(1), 156–184. https://doi.org/10.18608/jla.2015.21.8

Ren, Z., Ning, X., & Rangwala, H. (2017). Grade prediction with temporal course-wise influence. In *Proceedings of the 10th International Conference on Educational Data Mining* (pp. 48–55), Wuhan, China. International Educational Data Mining Society. https://arxiv.org/pdf/1709.05433.pdf

Right to Know. (2019). *Report card for education transparency and access*. http://righttoknowco.org/report-cards/

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics*, *2*(1), 7–12. https://doi.org/10.18608/jla.2015.21.2

Sabourin, J., Mott, B., & Lester, J. (2011). Modeling learner affect with theoretically grounded dynamic Bayesian networks. In S. D'Mello (Ed.), *Proceedings of the Fourth International Conference on Affective Computing and Intelligent Interaction* (pp. 286–295). Springer.

Sao Pedro, M., Jiang, Y., Paquette, L., & Baker, R. S. (2014). Identifying transfer of inquiry skills across physical science simulations using educational data mining. In *Proceedings of the 11th International Conference of the Learning Sciences* (pp. 222–229), Boulder, CO. International Society of the Learning Sciences. https://pdfs.semanticscholar.org/0ca7/cfc8e4575225162dd02c9ff64372158b030b.pdf

Scheihing, E., Vernier, M., Guerra, J., Born, J., & Carcamo, L. (2018). Understanding the role of micro-blogging in B-learning activities: Kelluwen experiences in Chilean public schools. *IEEE Transactions on Learning Technologies*, *11*(3), 280–293. https://doi.org/10.1109/TLT.2017.2714163

Schimke, A. (2019, September 10). *How are Colorado's kindergarteners doing? With hidden data, it's hard to tell*. https://chalkbeat.org/posts/co/2019/09/10/how-are-colorados-kindergarteners-doing-with-hidden-data-its-hard-to-tell/

Scott-Clayton, J. (2015). The shapeless river: Does a lack of structure inhibit students' progress at community colleges? In B. L. Castleman, S. Baum, & S. Schwartz (Eds.), *Decision making for student success* (pp. 114–135). Routledge.

Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics*, *2*(1), 13–48. https://doi.org/10.18608/jla.2015.21.3

Sekiya, T., Matsuda, Y., & Yamaguchi, K. (2015). Curriculum analysis of CS departments based on CS2013 by simplified, supervised LDA. In *Proceedings of the Fifth International Conference on Learning Analytics and Knowledge* (pp. 330–339). Association for Computing Machinery. https://doi.org/10.1145/2723576.2723594

Shen, S., & Chi, M. (2016). Aim low: Correlation-based feature selection for model-based reinforcement learning. In *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 507–512), Raleigh, NC. International Educational Data Mining Society. https://pdfs.semanticscholar.org/0acc/f1cf5015eb7cb7059b1b977991c1a9249e1a.pdf

Slater, S., Baker, R., Ocumpaugh, J., Inventado, P., Scupelli, P., & Heffernan, N. (2016). Semantic features of math problems: Relationships to student learning and engagement. In *Proceedings of the 9th International Conference on Educational Data Mining* (pp. 223–230), Raleigh, NC. International Educational Data Mining Society. https://pdfs.semanticscholar.org/df3e/9da476337def1021b12e03ab9959f9fd43d7.pdf

Strauss, V. (2019, September 11). Is New York State about to gut its student data privacy law? *Washington Post*. Retrieved from https://washingtonpost.com/education/2019/09/11/is-new-york-state-about-gut-its-student-data-privacy-law/

Sweeney, M., Lester, J., Rangwala, H., & Johri, A. (2016). Next-term student performance prediction: A recommender systems approach. *Journal of Educational Data Mining*, *8*(1), 22–51. https://doi.org/10.5281/zenodo.3554603

Tóth, K., Rölke, H., Greiff, S., & Wüstenberg, S. (2014). Discovering students' complex problem solving strategies in educational assessment. In *Proceedings of the 7th International Conference on Educational Data Mining*, London, UK. https://pdfs.semanticscholar.org/1baf/8106de777bd369cbc7eb4e1d8c6689aed420.pdf

Wang, D., Olson, J. S., Zhang, J., Nguyen, T., & Olson, G. M. (2015). DocuViz: Visualizing collaborative writing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1865–1874). Association for Computing Machinery. https://doi.org/10.1145/2702123.2702517

Wang, X., Yang, D., Wen, M., Koedinger, K., & Rosé, C. P. (2015). Investigating how student's cognitive behavior in MOOC discussion forums affect learning gains. In *Proceedings of the 8th International Conference on Educational Data Mining*. International Educational Data Mining Society. https://pdfs.semanticscholar.org/21af/f1b2e4f-fe5fbf86f3c4648315474ce271f1e.pdf

Wang, Z., Lan, A. S., Nie, W., Waters, A. E., Grimaldi, P. J., & Baraniuk, R. G. (2018). QG-net: A data-driven question generation model for educational content. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3231644.3231654

Ward, J. S., & Barker, A. (2013). *Undefined by data: A survey of big data definitions*. https://arxiv.org/pdf/1309.5821.pdf

Wen, M., Yang, D., & Rosé, C. P. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In *Proceedings of the 7th International Conference on Educational Data Mining* (pp. 130–137), London, England. International Educational Data Mining Society. https://pdfs.semanticscholar.org/3190/0b62fabf7da87573e93e473dd72cc68f24fa.pdf

Whitehill, J., Williams, J. J., Lopez, G., Coleman, C. A., & Reich, J. (2015). Beyond prediction: First steps toward automatic intervention in MOOC student stopout. In *Proceedings of the 8th International Conference on Educational Data Mining*, Madrid, Spain. International Educational Data Mining Society. http://dx.doi.org/10.2139/ssrn.2611750

Yang, H., & Meinel, C. (2014). Content based lecture video retrieval using speech and video text information. *IEEE Transactions on Learning Technologies*, *7*(2), 142–154. https://doi.org/10.1109/TLT.2014.2307305

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393

Yeung, C.-K., & Yeung, D.-Y. (2018). Addressing two problems in deep knowledge tracing via prediction-consistent regularization. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale* (pp. 1–10). Association for Computing Machinery. https://doi.org/10.1145/3231644.3231647

Zhang, L., & Rangwala, H. (2018). Early identification of at-risk students using iterative logistic regression. In C. Penstein Rosé, M. Mavrikis, R. Martínez-Maldonado, K. Porayska-Pomsta, H. U. Hoppe, B. McLaren, R. Luckin, & B. du Boulay (Eds.), *Artificial intelligence in education* (Vol. 10947, pp. 613–626). Association for Computing Machinery.