Chapter 1

Introduction

The daily activities of schools and universities—from taking attendance to assessing students—can leave a trail of data that, under the right conditions, can be used to explore teaching and learning like never before. Until recently, though, researchers had to choose between collecting rich data on a small number of individuals or amassing less rich data for larger numbers of individuals. And in both cases, collecting data on the same individuals over time required significant costs and complexities. For activities that take place in digital learning environments like games, learning management systems, and intelligent tutoring systems, surprisingly rich data can be collected on dizzyingly large numbers of learners over time. While opportunities to collect and analyze new forms of data increase every day, critical challenges need to be overcome in order to use these data to improve teaching and learning.

Along with new forms of data, such as system log data (i.e., records of users' interactions with a digital learning environment), familiar forms like text, audio, and video are becoming increasingly open to in-depth analysis—at scale—through machine learning and artificial intelligence. These newly found and newly analyzable data are often described as "big data" both inside and outside of education. Over the past decade, analyzing educational big data has largely occurred in research labs at universities, technology companies, and non-profit research institutes, and this basic research, with few exceptions, has yet to diffuse widely or to fundamentally change teaching and learning (Baker, 2016; Martin & Sherin, 2013). Where there have been successes, such as with the ASSISTments platform (Roschelle, Feng, Murphy, & Mason, 2016) and in examples described later on in this chapter, new forms of data and new analytical techniques have been grounded in problems facing practitioners and used to develop and assess potential changes related to those problems.

As some have argued, improving teaching and learning at scale will require new ways of organizing the work of educational research (Bryk, Gomez, Grunow, & LeMahieu, 2015). Starting around the same time as educational data mining and learning analytics—some of the most recognizable fields in what may be termed data-intensive research—an approach to conducting educational research referred to as research-practice partnerships was taking shape (Coburn, Penuel, & Geil, 2013). While the idea of forming partnerships is not new, frustrations with the status quo, a critical mass of success stories, and new funding opportunities have coalesced into an overarching approach where researchers work on pressing problems of practice in an iterative and collaborative fashion with practitioners (Penuel & Gallagher, 2017). In many ways, researchers working under the banner of research-practice partnerships have found a way to directly impact teaching and learning—by working directly with teachers and learners. While a disarmingly simple idea, this approach has profound implications both for *who* participates in the work of improving learning environments and for *how* that work is carried out.

In this book, we describe multiple efforts to use data-intensive research methods to improve teaching and learning. In particular, we highlight the important role that partnerships between researchers and practitioners can play in activating educational big data as a resource for improvement. Through the lens of what we refer to as *collaborative data-intensive improvement* (CDI), we aim to make explicit the ways in which educational researchers can engage in longer-term partnerships with the goal of not just understanding learning but also of improving outcomes in real-world learning environments. Doing this well, we believe, will require a fundamental rethinking of how data are used for research and improving practice.

Data-Intensive Research in Education

This book offers an introduction to the developing fields of educational data mining and learning analytics by describing goals, methods, and examples. In outlining the past, present, and potential future for these fields, throughout this book, we focus our descriptions on using data and complex data analyses to improve learning experiences and educational outcomes. We illustrate this potential with firsthand examples that span multiple academic content areas, learning environments, and learner types. We provide examples of decision making at the classroom, school, and education system levels taken from schools, universities, and community colleges.

Along with examples from our own work, we will describe how other researchers have employed educational data mining and learning analytics to address problems that originate in one form or another from the front lines of teaching and learning. In describing multiple examples and analytical approaches, we will highlight potential benefits and costs associated with each. The reader should know, however, that we are not attempting to provide a balanced treatment of all approaches. Our emphasis will be on *collaborative* data-intensive research approaches that prioritize shaping practical improvements over advancing analytic methods. While we will not restrict our coverage solely to collaborative data analysis approaches, they will be our lens for choosing what to highlight in a rapidly changing landscape. We hope that both researchers and practitioners will find this lens useful in making sense of new sources of education data, new analytic techniques, and new opportunities to form partnerships.

The Challenge of Jargon

One challenge facing newcomers to the field of data-intensive research is the wave of jargon they are likely to encounter. Already, in the first few pages of this book, we have referred to educational data mining, learning analytics, system log data, and big data. In an attempt to keep jargon to a manageable level, we have made explicit choices about the terminology we use in this volume, recognizing that some key details, distinctions, and research histories will be lost in this translation.

Before progressing further, we would like to orient the reader to a few key terms: educational data mining, learning analytics, data-intensive research, and educational data scientist.

Educational data mining and *learning analytics* represent distinct fields that have a high degree of overlap (Siemens & Baker, 2012). For simplicity, and to contrast these fields with other research traditions, we will refer to both of them as examples of *data-intensive research in education*. The additional fields that we want to integrate into learning analytics and educational data mining include studies of data use in schools (e.g., data-driven decision making) and collaborative research approaches (e.g., design-based implementation research and improvement science). As we will describe in Chapter 5, these additional fields are important both to the past and to the present of data-intensive research in the same ways that learning analytics and educational data mining are.

Educational data mining, which predates the field of learning analytics, largely concentrates on using machine learning techniques to identify patterns within large educational datasets, often from specific digital learning environments like intelligent tutoring systems. Oftentimes, these same technologies are what deliver interventions aimed at improving learning. Learning analytics, on the other hand, tends to focus less on machine learning techniques and more on statistical and visualization approaches, whereby interventions aimed at improving learning are delivered as much by an individual as a technology. As Baker and Inventado (2014) point out, the differences between these two fields grew out of different interests and backgrounds of the researchers in the two areas, and do not reflect any fundamentally opposing beliefs about how people learn. They agree on the assumption that data collected at scale and analyzed with rigorous methods will help arbitrate between different theories and proposed practices (Bienkowski, Feng, & Means, 2012).

Data-intensive research "involves data resources that are beyond the storage requirements, computational intensiveness, or complexity that is currently typical of the research field" (Dede, 2015, p. 2). The field of education more generally is gradually expanding its data repertoire to include data from digital learning environments and from increasingly sophisticated administrative data systems. In addition, other familiar forms of data, such as video and audio files, can now be explored at scale with greater speed. Therefore, we use the term data-intensive research to integrate these developing examples as well as those stemming from educational data mining and learning analytics.

An *educational data scientist* is someone who practices data-intensive research in education. The term "data scientist" is expansive and touches on multiple knowledge, skills, and abilities (see O'Neil & Schutt, 2013). Anyone who uses data-intensive research methods is often referred to as a data scientist. And while data science has become a hot new career (Ferenstein, 2016), the knowledge, skills, and abilities needed to perform this role are often ill-defined, especially in education (Piety, Hickey, & Bishop, 2014). Generally speaking, a data scientist is an individual with some combination of computer science skills, a background in statistics and mathematics, and relevant domain expertise (O'Neil & Schutt, 2013). Agasisti and Bowers (in press) define an educational data scientist as an individual who has "the technical skills to collect, analyze, and use quantitative data, and at the same time the managerial and communication skills to interact with decision-makers and managers at the school level to individuate good ways of using information in the practical way of improving practices and initiatives" (p. 6). In the coming chapters, we elaborate on these descriptions and make the case that an educational data scientist is someone who clarifies how data-intensive research methods can be used to address questions of importance to educators, carry out the actual analyses, and help develop and refine ideas for improvement.

Focus of the Book

Given the continuing proliferation of data and the increased sophistication of data-intensive research techniques, now is a good time to take stock of data-intensive research in education, articulate fruitful directions for advancing the field, and provide an onramp for newcomers. In working to achieve this ambitious and multifaceted aim, it is important to clarify what this book will and will not deliver. First, this book is not a how-to guide on data-intensive research methods in education. The interested reader can explore a growing number of learning analytics focused Massive Online Open Courses (MOOCs) for this purpose, such as Ryan Baker's *Big Data and Education*, Tim McKay's *Practical Learning Analytics*, and the University of Texas at Arlington's upcoming MicroMasters on Learning Analytics. In addition to educational applications of analytics, a researcher or data scientist, at some point, will need to group rows of data and apply a function, such as identifying the average amount of time a student spent in a digital learning environment across multiple sessions. Depending upon one's chosen software package, without too much difficulty, one could use a search engine to identify a serviceable answer. Less searchable are strategies for identifying sources of data in the first place and knowing how to work with practitioners to apply the right analytical technique to the right data and how to structure a meeting where researchers and practitioners come together to interpret and draw implications from a data-intensive analysis. In many ways, that is what this book is about.

This book is also not a standard course in educational research design or a program in educational leadership, though it does include elements and insights from these fields. It presents some fundamental research and leadership concepts as they relate to each other and to the goal of using data-intensive research to improve education outcomes. We seek to equip readers with an understanding of methods to enable clearer thinking about how new sources of data and new analytical techniques could help them create more desirable outcomes for students.

Examples of Data-Intensive Improvement

When Romero and Ventura (2007) surveyed the data mining literature for education applications published between 1995 and 2005, they found only two articles published before 2000. In contrast, by 2016, a Google Scholar search returned over *one million articles* on this topic. And educational applications of data-intensive research have moved beyond scholarly publications to capture the public's imagination through popular press coverage such as a recent *New York Times* article, "Will You Graduate? Ask Big Data" (Treaster, 2017). In the following sections, we describe three diverse examples to introduce some of the possibilities.

Measuring Chronic Absenteeism and Its Causes

School districts have always kept data on their students, but it used to be hard to access or to organize the data in a way that would shed light on educational issues. For example, a school district would have a record system showing the number of students in attendance each day and would routinely compute the average daily attendance for the school year. Schools with attendance average daily rates over 90 percent generally believed they were doing very well on this metric. But most schools, districts, and states did not have the capability to look at attendance patterns for *individual students* over multiple years or to relate students' attendance patterns to their educational outcomes (Balfanz & Byrnes, 2012). Without such a longitudinal student-level dataset, schools were missing the story of what has come to be called "chronic absenteeism"—missing 10 percent or more of school days in an academic year.

The importance of attending school has long been recognized, but until recently we lacked the ability to quantify the impact of chronic absenteeism on educational outcomes and hence any basis for saying what level of absenteeism should be cause for concern. Increased computing capacity, improved tools for bringing together data from different data systems, and the use of unique, statewide student identification numbers permitting linking multiple student-level datasets have enabled exploration of the issue of absenteeism in states and districts.

Analysis of data from Chicago Public Schools (CPS) by researchers from the University of Chicago Consortium on School Research, for example, found that missing 10 or more days of school during the year, whether excused or not, was a stronger predictor of school failure than low test scores at the end of the prior school year. Analysis of the CPS data also showed that ninth graders with high test scores who missed two or more weeks of school were more likely to fail than students with low test scores who were absent five or fewer days (Allensworth & Easton, 2007). An issue brief from the University of Chicago's To&Through project indicated that each week a student is absent during a semester of ninth grade is associated with a 20 percent decline in the probability of earning a high school diploma. After becoming aware of the data on chronic absenteeism and its correlates, CPS began implementing a number of programs to address chronic absenteeism. One strategy involved improving the accuracy and availability of individual students' attendance records so that teachers and school leaders would be motivated to examine them on a weekly basis in order to identify students in need of intervention. Another strategy involved creating a culture of collective responsibility around attendance. Some schools started talking about the importance of attendance at school assemblies and posted attendance charts in school hallways. The combination of these and other approaches over the past decade have led to a 17 percent increase in high school graduation rates (To&Through Project, no date).

Using Learning Analytics to Improve Digital Learning Systems

System log data from digital learning environments are particularly promising because they can capture *who* did *what* and *when*. Researchers and educational data scientists can explore this kind of data to look at the sequences of actions taken by individual learners, greatly expanding the potential to examine detailed learning activity data at a massive scale in order to glean insights into the *processes* of learning. Being able to go beyond analysis of outcomes to delve into learning processes opens up significant opportunities for improving both digital and face-to-face learning environments.

Since the 1980s, Carnegie Mellon University (CMU) has pioneered the design, development, and evaluation of digital learning systems that employ learning theory and artificial intelligence to adapt to the responses of individual learners (Koedinger & Corbett, 2006). More recently, with the availability of increased data storage and analysis capabilities, researchers at CMU began applying a variety of machine learning and statistical techniques to the data produced when students use their tutoring systems in order to derive insights into how to improve those tutoring systems (Koedinger, Stamper, McLaughlin, & Nixon, 2013).

A hallmark of the tutoring systems developed at CMU is that they are based on a detailed cognitive analysis of the knowledge and skill components needed in the domain being studied. Each problem presented in the tutoring system was designed to assess one or more knowledge components. One of the types of data researchers extract from the tutoring system's log files is whether the learner made an error or answered correctly each problem involving a given knowledge component. Ken Koedinger and Elizabeth McLaughlin of CMU leveraged this kind of data in a recent study in which middle school students solved large numbers of beginning algebra problems online, including the three problem types shown in Table 1.1. The target proficiency in this study was being able to solve two-step story problems, such as the one shown in the left-hand column. The researchers wanted to figure out what kind of practice would best support students in acquiring this skill.

Table 1.1 Story Problem Types Studied by Koedinger and McLaughlin

	Problem Type	
2-step Story Problem	*1-step Story Problem*	*Substitution Problem*
Ms. Lindquist is a math teacher. Ms. Lindquist teaches 62 girls. Ms. Lindquist teaches *f fewer boys* than girls. Write an expression for how many students Ms. Lindquist teaches.	Ms. Lindquist is a math teacher. Ms. Lindquist teaches 62 girls. Ms. Lindquist teaches *b boys*. Write an expression for how many students Ms. Lindquist teaches.	Substitute 62-f for b in 62+b Write the resulting expression.
Answer: 62+62-f	Answer: 62+b	Answer: 62+62-f

Source: Koedinger and McLaughlin (2016).

Many instructional designers and educators would hypothesize that practicing one-step problems to mastery would be the best foundation for moving to the harder two-step problems. Although this makes intuitive sense, a major theoretical assumption in the CMU work is that difficulty levels predict transfer because both are a function of the same underlying required knowledge components. When the Carnegie Mellon team analyzed data from the log files for these three types of problems, they found that students had more difficulty with substitution problems, like that in the third column, than they did with one-step story problems. For this reason, Koedinger and McLaughlin predicted that it would be more beneficial to practice symbolizing algebraic terms in the substitution problems than to practice one-step story problems.

Using a web-based tutoring system, the researchers randomly assigned 711 middle school math students to either substitution practice or one-step story problem practice in preparation for two-step story problems. Findings supported the researchers' hypothesis that production of symbolic representations was the key prerequisite for learning to solve the two-step algebra problems. Prior practice on substitution problems based on the cognitive model generated from this data-driven approach inspired an intervention that subsequent experimental testing showed would enhance learning on the target skill of two-step algebra word problems.

The researchers interpret this finding as support for their assumption that task difficulty data can be used as a proxy for skill transfer data. They point out the practical significance of this finding: Direct testing of the transfer of skills from one type of problem to another requires setting up an experiment to test performance on task B with and without prior practice on task A. Generating task difficulty parameters automatically through data mining can provide the input needed for cognitive models so that instructional design and development work can proceed more quickly and more ethically (Koedinger & McLaughlin, 2016).

Identifying College Students at Risk of Dropping Out

Our third example returns to the challenge of identifying students at risk of leaving school, but in this case at the college level. Earning a college degree has major consequences for employability and lifetime earning (Pascarella & Terenzini, 2005). Thanks to the wide range of higher education options in the U.S., including institutions with open admissions, increasing proportions of young people from all backgrounds start some kind of college program. But *completing* a college program with a degree or industry-recognized credential is something different. For students entering college for the first time in 2009, for example, only 53 percent earned a bachelor's degree by 2015, six years later.

As state and federal governments have increased their scrutiny of completion rates for individual colleges, those institutions have become acutely aware of the need to increase the proportion of their students who are retained from year to year and actually leave with a degree. Colleges and universities have turned to data-intensive research techniques to help them identify students who are at risk of failing to complete a course or program of study. Measuring graduation rates requires connecting the academic records from different terms for each individual student to measure whether that student persisted from one term to the next. By combining data from admissions applications and transcripts with data on performance in a particular course, analysts found they could identify groups of students at risk so that those students' instructors or academic advisors could work with them to avoid course failure and dropping out (Hanover Research, 2014).

Tim Renick, Vice President for Enrollment and Student Success at Georgia State University, describes a well-known case of using data-intensive approaches to enhance college completion rates (Renick, 2017). In 2003 this urban public university saw just 33 percent of white students who had enrolled as freshmen and just 22 percent of under-represented minorities who had enrolled as freshmen leave the college with a bachelor's degree. By 2017, Georgia State's degree completion rate had risen to 65 percent for both groups of students, making Georgia State the only public university in the nation where the completion rate for under-represented students is equivalent to that for white students.

Georgia State implemented multiple changes in its practices and interventions with students at risk to achieve these results (Kurzweil & Wu, 2015), but a key enabler was a collaboration with EAB (formerly the Education Advisory Board). EAB helped Georgia State comb through 10 years of student data records—over 2.5 million course grades. These analyses provided insights such as the fact that prospective political science majors who got an A or B in their first political science course had a 75 percent probability of graduating on time, while those who got a C had only a 25 percent probability of doing so (evoLLution, 2016). The university had been doing nothing to follow up with students who earned Cs in their gateway courses because a C grade is adequate to earn the course credit toward graduation. Georgia State hired more academic advisors in order to act more promptly on information identified by an analysis.

The Graduation and Progression System (GPS) academic advising dashboard developed by Georgia State and EAB displays real-time analyses of students' academic progress and the implications of certain decisions, such as taking courses out of the usual sequence.

The GPS displays results from a system that tracks students for 800 different alerts that can trigger action:

> Now, every day the system searches all of our student-information systems for evidence of any of these 800 things. Did a student register

> for the wrong course? Did they do poorly in a prerequisite course? Are they in a major that does not fit their ability? When an alert goes off, an advisor proactively reaches out to the student, typically within 48 hours.
>
> (evoLLution, 2016)

Higher education institutions are also starting to combine the relatively stable information from academic records with more timely information from their campus learning management systems. Learning management systems (LMSs) are online systems that support instructors in delivering course content and assessments; many LMSs include interactive features such as discussion boards. Measures such as the number of days on which a student logs in to the LMS compared with other students in their class, scores earned on interim assessments within the LMS, engagement with course materials, and participation on discussion boards, all measured relative to other students in the same course, can in certain circumstances be used to predict likelihood of completing the course. Combining LMS data with other types of data, the firm Civitas Learning has helped several of its client institutions identify individuals among their high-GPA students who were showing signs of disengagement with college. These high-achieving disengaged students had tended to fall through the cracks because they did not have obvious markers of course failure in their academic records.

Common across all the examples cited previously is the use of data-intensive research methods for identifying and working to improve educational processes. In all cases, data helped identify an opportunity to improve but the data didn't solve the problem—that was up to teams of people in each education institution.

Why Engage in Collaborative Data-Intensive Improvement?

In the examples cited previously, the work of identifying problems to solve, collecting and analyzing data, and deriving implications is prototypical of the work of researchers and educational data scientists engaging in a style of inquiry conducted in concert with educators that we refer to as *collaborative data-intensive improvement* (CDI).

As the use of digital learning environments in schools increases and more and more data are captured in administrative data systems, researchers and data scientists who can support CDI may increasingly be called upon to not only extract meaning from data but also to structure specific activities before and after developing data products. These before and after activities are critical, as they help partnerships translate what is learned from a data-intensive analysis into specific actions that can be used to solve local problems of practice (Krumm, Waddington, Teasley, & Lonn, 2014).

Along with the growing evidence for the benefit of combining data-intensive research with specific efforts to improve teaching and learning, three trends give us optimism that the time is ripe for engaging in data-intensive research in education, and in particular, CDI:

- ***Lessons learned from the data driven decision-making movement.*** For decades, schools have been pressed to use data to drive their instructional and organizational decision making. Multiple scholars have examined what worked and what didn't from this period, and as a field, we are moving beyond viewing data as inherently actionable or as a self-activating resource.
- ***Increased role and importance of research–practice partnerships.*** Both private philanthropies and federal agencies are supporting this trend by providing funding for collaborations between researchers and practitioners. Pioneering efforts from organizations, such as the Carnegie Foundation for the Advancement of Teaching, are providing useful models for how these partnerships can work (Coburn & Penuel, 2016).
- ***The availability of data and the need to interpret them responsibly.*** Data of increasing size and variety are available as never before. With this growing resource will come a need to structure appropriate analyses, draw appropriate conclusions, and structure follow-on activities.

Engaging in CDI opens up unique possibilities for researchers and educational data scientists; education leaders and practitioners; and technology developers.

For researchers and educational data scientists who want to see their work improve the quality of education and the equity of opportunities that students receive, collaborative forms of data-intensive research offer opportunities to directly experience and participate in the improvement process. Researchers are accustomed to publishing their analyses and research conclusions in technical reports and scholarly journal articles, which are often not read by the education decision makers and practitioners responsible for the educational experiences that students actually receive. Even when a study does get wide publicity in the general press or in venues where educators gather information, such as their professional conferences or periodicals, a research report is not self-explanatory; understanding how to apply an insight from research to a new context is challenging for researchers and practitioners alike. By directly engaging with practitioners in a partnership, researchers have the opportunity to see their work put to use in real learning environments in ways they believe are well reasoned and likely to be successful.

For education leaders, engaging in data-intensive work with researchers offers an opportunity to increase the likelihood of ameliorating an important problem of practice by applying a systematic set of tools and approaches and enlisting additional intellectual resources in the form of researchers and data scientists. The magnitude of improvement that is possible has been demonstrated by the Carnegie Math Pathways from the Carnegie Foundation for the Advancement of Teaching, which we will describe throughout this book. In some cases, university systems and school systems eager to apply data-intensive approaches to their improvement efforts are funding the work of their external collaborators directly (Treaster, 2017), but in other cases researchers have their own funding to support their participation (UT Arlington News Center, 2014).

For teachers and instructors, CDI complements a commitment to the scholarship of teaching and learning. This form of scholarship entails reflective inquiry into student learning in specific academic domains and seeks to generate insights that improve teaching and thereby enhance student learning (Hutchings, Taylor Huber, & Ciccone, 2011). Lee Shulman, one of the early advocates for this form of inquiry, offers three rationales for this kind of work that are equally applicable to CDI (Shulman, 2000). First, there is professionalism, which Shulman describes as the "inherent obligation" entailed in being a professional educator and in representing the discipline one teaches. Second, there is the pragmatic rationale: An educator should strive to make sure that his or her work is constantly improving and enabling students to meet their learning goals. Finally, there is the need to be able to demonstrate to external authorities such as administrators, school boards, and accrediting agencies that one's teaching is adding value for students and improving over time. CDI can enhance the scholarship of teaching and learning by convening collaborators with diverse expertise and an expanded set of methods.

For learning technology developers, participation in the types of partnerships and collaborations described in this book can be used to expand their internal capacity for research and analytics as well as gathering new insights into issues surrounding the implementation of their products. Collaborating with researchers and practitioners can help technology developers gain a fuller understanding of the things that are important to teachers—their potential customers—and to supporting student learning. Moreover, if they design their learning system with the idea of being able to provide data collection and storage infrastructure that can be later analyzed efficiently, they will be better prepared to drive future enhancements of their products. We have found that a surprising number of learning technology products are developed and marketed widely without the capability to capture the kind of data that can inform teaching and learning. If developers understand how data captured by their technologies can be used to improve teaching and learning, they can be better equipped to collect and store data that can support continuous improvement of their products and how they are used.

Contents of This Book

This book seeks to support both researchers, practitioners, and developers in applying data-intensive research methods to improve learning environments. Our goal is to offer scaffolds that a team can use to develop a research–practice partnership and use data with the rigor needed to make meaningful progress.

This introduction to our approach for merging data-intensive research, improvement science, and educational research will be followed by a description of the kinds of data that are available for use within research–practice partnerships in Chapter 2. Chapter 3 then introduces analytical techniques used in data-intensive research projects at an introductory level. In Chapter 4, we discuss issues of data privacy and security as well as approaches for using student data for research and improvement purposes. In Chapter 5, we describe the influences that have fostered an increased reliance on data and evidence in educational decision making and various conceptions of how researchers and education practitioners should work together in greater detail. These traditions influenced our own research and provided a foundation for our model of CDI. Using two cases, Chapter 6 presents our CDI model and discusses key assumptions of the model. Chapter 7 provides a deep dive into CDI practices and tools, presenting five phases for implementing this kind of work. Finally, we conclude in Chapter 8 with a summary of some of the key things we have learned from our work and the work of others and an explication of trends that are likely to shape future applications of data-intensive research in education.

References

Agasisti, T., & Bowers, A. J. (in press). Data analytics and decision-making in education: Towards the educational data scientist as a key actor in schools and higher education institutions. In G. Johnes, J. Johnes, T. Agasisti, & L. López-Torres (Eds.), *Handbook on the economics of education*. Cheltenham, UK: Edward Elgar Publishing.

Allensworth, E. M., & Easton, J. Q. (2007). *What matters for staying on track and graduating in Chicago Public Schools*. Chicago, IL: University of Chicago Consortium on Chicago School Research.

Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education*, *26*(2), 600–614. doi:10.1007/s40593-016-0105-0

Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In J. A. Larusson & B. White (Eds.), *Learning analytics* (pp. 61–75). New York: Springer. doi:10.1007/978-1-4614-3305-7_4

Balfanz, R., & Byrnes, V. (2012). *Chronic absenteeism: Summarizing what we know from nationally available data*. Baltimore: Johns Hopkins University Center for Social Organization of Schools.

Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*. Washington, DC: U.S. Department of Education.

Bryk, A. S., Gomez, L. M., Grunow, A., & LeMahieu, P. G. (2015). *Learning to improve: How America's schools can get better at getting better*. Cambridge, MA: Harvard Education Press.

Coburn, C. E., & Penuel, W. R. (2016). Research-practice partnerships in education: Outcomes, dynamics, and open questions. *Educational Researcher*, *45*(1), 48–54.

Coburn, C. E., Penuel, W. R., & Geil, K. E. (2013). *Research-practice partnerships: A strategy for leveraging research for educational improvement in school districts*. New York: William T. Grant Foundation.

Dede, C. J. (Ed.). (2015). *Data-intensive research in education: Current work and next steps*. Computing Research Association. Retrieved from http://cra.org/wp-content/uploads/2015/10/CRAEducationReport2015.pdf

evoLLution. (2016). The scale of change in higher education: Using technology and impacting student success. Retrieved June 27, 2017, from https://evolllution.com/technology/tech-tools-and-resources/the-scale-of-change-in-higher-education-using-technology-and-impacting-student-success/

Ferenstein, G. (2016, January 20). *Why 'data scientist' is the best job to pursue in 2016*. Forbes. Retrieved October 1, 2016, from www.forbes.com/sites/gregoryferenstein/2016/01/20/report-why-data-scientist-is-the-best-job-to-pursue-in-2016/#7a1432f45f4b

Hanover Research. (2014). *Early alert systems in higher education*. Washington, DC: Author.

Hutchings, P., Taylor Huber, M., & Ciccone, A. (2011). *The scholarship of teaching and learning reconsidered: Institutional integration and impact*. San Francisco: Jossey-Bass.

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

Koedinger, K. R., & McLaughlin, E. A. (2016). Closing the loop with quantitative cognitive task analysis. Presented at the *9th International Conference on Educational Data Mining*. Raleigh, NC. Retrieved from www.educationaldatamining.org/EDM2016/proceedings/paper_152.pdf

Koedinger, K. R., Stamper, J. C., McLaughlin, E. A., & Nixon, T. (2013). Using data-driven discovery of better student models to improve student learning. In *Proceedings of the 16th international conference on artificial intelligence in education* (pp. 421–430). Phoenix, AZ.

Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). Using learning analytics to support academic advising in undergraduate engineering education.

In J. A. Larusson & B. White (Eds.). *Learning analytics: From research to practice* (pp. 103-119). New York: Springer.

Kurzweil, M., & Wu, D. D. (2015). *Building a pathway to student success at Georgia State University: A case study*. New York: ITHAKA S+R.

Martin, T., & Sherin, B. (2013). Learning analytics and computational techniques for detecting and evaluating patterns in learning: An introduction to the special issue. *Journal of the Learning Sciences*, *22*(4), 511–520.

O'Neil, C., & Schutt, R. (2013). *Doing data science: Straight talk from the frontline*. Sebastopol, CA: O'Reilly Media.

Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students, vol. 2: A third decade of research*. San Francisco: Jossey-Bass.

Penuel, W. R., & Gallagher, D. J. (2017). *Creating research-practice partnerships in education*. Cambridge, MA: Harvard Education Press.

Piety, P. J., Hickey, D. T., & Bishop, M. J. (2014). Educational data sciences: Framing emergent practices for analytics of learning, organizations, and systems. In *Proceedings of the 4th international conference on learning analytics and knowledge* (Indianapolis, IN) (pp. 193–202). New York: ACM. doi:10.1145/2567574.2567582

Renick, T. (2017, May). *Keynote address for EdTech efficacy research academic symposium sponsored by University of Virginia Curry School of Education and Digital Promise*. Washington, DC.

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

Roschelle, J., Feng, M., Murphy, R. F., & Mason, C. A. (2016). Online mathematics homework increases student achievement. *AERA Open*, *2*(4), 1–12. doi:10.1177/2332858416673968

Shulman, L. S. (2000). From Minsk to Pinsk: Why a scholarship of teaching and learning? *The Journal of Scholarship of Teaching and Learning*, *1*(1), 48–53.

Siemens, G., & Baker, R. S. J. D. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In S. B. Shum, D. Gasevic, & R. Ferguson (Eds.), *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252–254). New York: ACM. doi:10.1145/2330601.2330661

To&Through Project. (no date). To&Through issue brief: Attendance. Retrieved June 28, 2017, from https://toandthrough.uchicago.edu/resources

Treaster, J. B. (2017, February 2). Will you graduate? Ask big data. *New York Times*. Retrieved from www.nytimes.com/2017/02/02/education/edlife/will-you-graduate-ask-big-data.html?_r=0

UT Arlington News Center. (2014, November). UT Arlington to lead $1.6 million research project focused on digital learning. Retrieved June 27, 2017, from www.uta.edu/news/releases/2014/11/LINKLab-dLRN.php