TechTrends (2018) 62:37–45
DOI 10.1007/s11528-017-0226-9

ORIGINAL PAPER

Learning Analytics: at the Nexus of Big Data, Digital Innovation, and Social Justice in Education

Stephen J. Aguilar[1]

Published online: 10 October 2017

Abstract We are still designing educational experiences for the *average* student, and have room to improve. Learning analytics provides a way forward. This commentary describes how learning analytics-based applications are well positioned to meaningfully personalize the learning experience in diverse ways. In so doing, learning analytics has the potential to contribute to more equitable and socially just educational outcomes for students who might otherwise be seen through the lens of the average student. Utilizing big data, good design, and the input of the stakeholders, learning analytics techniques aim to develop applications for the sole purpose of reducing the classroom size to 1. Over time, these digital innovations will enable us to do away with a model of education that teaches toward the non-existent average student, replacing it with one that is more socially just—one that addresses the individual needs of every student.

Keywords Learning analytics · Social justice · Digital innovation · Educational technology · Personalized learning · Applications

> "*Only by being true to the full growth of all the individuals who make it up, can society by any chance be true to itself.*" (Dewy 1915)

Education has the potential to be a powerful mechanism for social change, upward mobility, and social equity—but only if it reaches everyone; or, as John Dewy writes, if it contributes to the "full growth" of all of society's individuals (Dewy 1915). Yet, there is tension between the aspirations of education and the realities associated with its implementation. Often, education is seen as a mechanism for upward mobility, yet the quality of students' educational experiences differ. Teacher quality, for example, has been shown to be both vital and variable, leading to differences in student outcomes (Darling-Hammond 2000). This is complicated by the fact that instructional practices do not always align with content standards (Polikoff 2012), undercutting the latter's potential to improve student achievement.

It is challenging to reach *every* student in a personalized way. As a result, students are organized into groups, e.g., grade levels, classrooms, and schools. Teachers and administrators who oversee students, then, must differentiate instruction to meet the needs of each learner. Differentiation is challenging, and stakeholders often compromise and teach to the *average* student. Smaller class sizes help, and they have been shown to increase the probability of promotion to the next grade (Gary-Bobo and Mahjoub 2013), benefit members of racial and ethnic minority groups (Finn and Achilles 1990), and have been associated with higher academic achievement for black students across multiple domains (Shin 2012). Smaller class sizes also improve grades in collegiate settings (Johnson 2010). "Small," however, is relative. Research suggests that an "optimal" class size does not always refer to a small one (Borland et al. 2005).

Smaller class sizes speak to an intuitive characteristic of education; namely, that it is easier to teach and learn when instruction is personalized. Personalized instruction is not a new idea. Fred Keller (1968, 1974) formalized a definition of it almost five decades ago. He dubbed his method the "Personalized System of Instruction," or PSI. It focused on making course content incremental, self-paced, mastery focused, iterative, and supported by others (Keller 1974). In

✉ Stephen J. Aguilar
aguilars@usc.edu

[1] University of Southern California, 3470 Trousdale Parkway, 600B Waite Phillips Hall, Los Angeles, CA 90089-4036, USA

one of his first syllabi to adopt the method, Keller writes: "This is a course through which you may move, from start to finish, at your own pace. You will not be held back by other students or forced to go ahead until you are ready." Importantly, the system relied on a team: "The *teaching staff* of your course will include proctors, assistants, and an instructor" (emphasis mine). Over time, it was shown to be effective (Kulik et al. 1979). The challenge, however, is that PSI requires vast amounts of human capital to scale.

Educational technologies have begun address this challenge by taking elements of personalized learning to scale. Though not necessarily based on PSI, cognitive and intelligent tutors (Anderson et al. 1985) have been developed for close to 20 years (Murray 1999). They rely on having computers do the personalization (Anderson et al. 1985), and have been shown to promote meta-cognitive strategies (Aleven and Koedinger 2002; Aleven et al. 2006). Unfortunately, even cognitive and intelligent tutors are not immune to failure, because their strength (i.e., scalability) is also a limitation. Unbeknownst to the teacher, students can engage in off-task behavior while using them, or can game the system for the sake of completing tasks quickly (Baker et al. 2006).

In what follows, I argue that *learning analytics* driven educational technologies—and the *big data* that underpin their development—bring us one step closer to eliminating teaching toward the average. Learning analytics-driven technologies are capable of providing learners and instructors alike with insights into learning processes while also detecting students who struggle or become disengaged, making it less likely for those students to fall through the cracks. The unique capabilities of learning analytics-driven educational technologies speak to their potential to be a positive force for social change, and they are poised to have a central role in the reduction of social inequity because they represent a unique, scalable learning technology that can be used to address *individual* students' learning needs.

Big Data and Learning Analytics

"Big Data" is distinct from traditional sources of data in four fundamental ways. First, it is more *voluminous*; it is estimated that 2.3 trillion gigabytes of data are generated every day. For context, data generated during 2 days in 2011 was larger than all of the data generated from the origin of civilization to 2003 (Chen et al. 2014). Second, there is a large variety of data sources, ranging from text generated through Facebook posts, to data gathered by wearables, to video content created on YouTube. Third, big data has a *velocity*, i.e., analysis can occur quickly because big data can "stream," or move from instruments to analysis quickly. Sensors in an automobile, for example, can detect and report tire pressure, fuel levels, and geolocation—all of which are measured constantly. Finally, big data can be unruly, leading to doubts concerning its veracity. Together, volume, variety, velocity, and veracity make up the four Vs of big data (IBM 2017).

Given the four Vs of big data, stakeholders use it in different ways, and each sector has responded to the four Vs of big data differently. In business, for example, there has been widespread use of big data (i.e., "business analytics") to inform decisions regarding supply chains (Trkman et al. 2010). In the healthcare industry, the analysis of big data collected by providers has led to improved use of resources and better decision making, saving an estimated 29,000 lives and reducing spending by an estimated 7 billion dollars (Raghupathi and Raghupathi 2014).

In education, *learning analytics* has emerged as the discipline associated with analyzing and reporting big data, though work in this area is relatively new. Fields dedicated to the development of learning analytics applications are rapidly expanding (Siemens 2013), and many programs now offer degrees in learning analytics research and practice. Indeed, the 2014 Horizon Report published jointly by the New Media Consortium and the EDUCAUSE Learning Initiative predicted that learning analytics—which provide "ways...to improve student engagement and provide a high-quality, personalized experience for learners"—will be soon be adopted widely (Johnson et al. 2014). Predicated on the availability of large amounts of student data (e.g., student demographic information from student information systems, student grades, and/or student behaviors in learning management systems), learning analytics is poised to transform the ways in which schools and universities leverage big data to address issues of retention and student success, thus moving away from the average student and towards addressing the needs of each student in a personalized and data-driven way. Students themselves, moreover, may soon have access to new tools that enable them to be more reflective of their own learning; learning analytics-driven educational technologies have the potential to unveil learning processes that were traditionally hidden, or not feasible to measure, such as patterns associated with online discussion behaviors (Wise et al. 2014). Notably, however, work in *student-facing* learning analytics is still relatively new (Bodily and Verbert 2017).

Learning analytics focuses primarily on analyzing and representing "data about learners in order to improve learning" (Clow 2013), and learning analytics users include students (Wise 2014) or intermediaries, such as academic advisors (Aguilar et al. 2014). Recently studied learning analytics topics include understanding students' behaviors in online learning systems (Brooks et al. 2014), predictive modeling of student outcomes (Teplovs et al. 2011) and learning analytics methodologies (Suthers and Rosen 2011).

What distinguishes learning analytics from traditional modes of academic feedback (e.g., grades, or comments on an essay) is its capacity to extract data from user interactions with the learning environment, analyze it, and "load" it onto applications that are used by those invested in the learning process (Lonn et al. 2013). The resulting feedback is designed to be both personal and provide information that helps users identify trends both within and across different students, courses, and/or student populations. This feedback also creates *actionable* information that informs academic decision making (Arnold and Pistilli 2012), influences students' academic motivation and approaches to learning (Lonn et al. 2014).

The Fallacy of the "Average"

Learning analytics, and its reliance on big data found in education settings, is important because it has the potential to move us away from the fallacy of designing toward the average. The average, i.e., the statistical mean of a given measurement, has historically been an ineffective tool for design. After World War II, the United States Air Force, for example, had a major problem on its hands: its pilots were having difficulty controlling the new jet planes that were being manufactured—some pilots would lose control of their aircraft in catastrophic ways and lose their lives in the process (Daniels 1952). Early investigations into the problem concluded that pilot error was most likely the cause, but further investigations determined that the design of the cockpit was at the root of the problem. Engineers tasked with redesigning it decided that they needed to calculate the physical dimensions for a new "average" pilot. Their previous cockpit design, after all, was based on pilot measurements that were done in the 1920s (Daniels 1952). This led to development of an ambitious plan for determining the proper dimensions for new cockpit designs. The protocol developed seemed simple and achievable: to build a cockpit for the average man (the use of "man" here is deliberate—women were not seen as potential pilots), the Air Force would measure over 140 physical dimensions of 4000 of pilots and take the average of those measurements. The results were intended to yield an "average" pilot—one whose physical dimensions would inform the design of new cockpits. This would hypothetically help engineers design a cockpit that could work for the majority of its pilots.

Unfortunately, this approach was unsuccessful. In his technical report written for the US Air Force, Gilbert Daniels (1952) showed that out of all the pilots measured, exactly *zero* fit into the average range of just *ten* of the dimensions measured. Daniels concluded that the average, for all its merits, was a woefully inadequate statistic to use in the design process:

> "While the use of average dimension is generally unsatisfactory even when only one dimension is being considered at a time, the inadequacy of the 'average man' method is compounded many times when more than one dimension is to be considered in a design problem." (Daniels 1952 pg. 2)

Daniels highlights an insidious problem in design, namely, if you design for an "average" person, not only are you designing for no one in particular, you are, in fact, designing for no one at all.

The "Average" Student

In education the problem is no less insidious. Rather than the average person, there have been debates about what the average student needs to succeed. This debate plays out at the most fundamental level of school design. After a set amount of instructional time (an academic year, for example), it is expected that students have learned what is appropriate for their grade level. Although one can take an average of students' needs and achievements, one cannot necessarily apply that average to any student *in particular*. Thus, while it can be said that the *average* American 9th grader has a working understanding of Algebra, various factors influence the truth of this statement. Nor can one average across schools, because this ignores important contextual factors such as class size, the "ethos" of the school, and the effects of staff (e.g., Lee and Bryk 1989).

Creating a personal learning experience falls onto teachers, who are given the monumental task of differentiating their instruction in an attempt to meet the needs of all of their students (Lawrence-Brown 2004; McTighe and Brown 2005). In an effort to help personalize instruction, educators have sometimes had to find creative ways to reduce class size, including "tag-team" teaching (Graue et al. 2007). These efforts, however, have had mixed results (e.g., Gamoran and Weinstein 1998). Indeed, the lack of variation in instruction (i.e., teaching to the average), leads to more variation in student achievement outcomes (Gusky 2007). It is understandable, then, that achievement gaps have emerged (Ladson-Billings 2006; J. Lee 2002). Unfortunately, they often disproportionally affect ethnic and racial minorities (Quinn 2015; Vanneman et al. 2011).

Work to reduce gaps based on ethnicity, race, and gender is ongoing, and has spanned over 30 years (see Lee 2002 for an overview of Black-white and Hispanic-White achievement gaps, or Curran and Kellogg 2016 for K-1 race/ethnicity gaps in STEM). Some solutions have included providing more service-learning opportunities for at-risk students (Scales et al. 2006), or have used an intervention designed to affirm their values (e.g., friendship) during a course (Miyake et al. 2010). Involving student input in the design and implementation of assessments has also been shown to reduce achievement gaps (Stiggins and Chappuis 2005). Personalized instruction has also helped to improve the achievement gap between lower- and upper-middle class students (Pennebaker et al. 2013).

Learning analytics-driven educational technologies will move us away from thinking about the "average" student, while simultaneously empowering those invested in education to focus on the individual needs of their students. These technologies have the potential to personalize learning, and are well positioned to contribute to socially just and equitable student outcomes because they do away with the notion of an "average" student. Instead, they allow teachers, administrators, and other stakeholders to personalize the learning experiences of students.

Moving beyond the "Average" Student

A defining characteristic of learning analytics is to utilize data in any form it comes in, and tie it to a particular student for the purpose of personalizing their learning experience. In the past, much of this work has occurred in digital learning environments, such as cognitive tutors (Aleven et al. 2006; e.g., Aleven and Koedinger 2002; Baker et al. 2006; Baker et al. 2010). Recently, the contexts where data is gathered from have widened; sometimes requiring the capability to merge data from both digital and "analogue" sources to study various types of learning environments. Miller et al. (2015), for example, used field observations of teachers' use of proactive remediation (i.e., helping struggling students at the moment of confusion) in order to build an algorithm for detecting remediation without the need for an in-person field observation. To do this, Miller et al. merged field observation data with log files from a digital mathematics curriculum designed for blended learning environments and were able to build a robust model that detected teachers' proactive remediation 60% of the time (Miller et al. 2015). Their results show the possibility of leveraging learning analytics techniques to better scale what were once resource-intensive interventions like field observations.

Similarly, *multimodal* learning analytics aggregate and analyze data from distinct sources (multiple modalities), and present insights via tools that collect "process" learning data that is generated *during* a learning activity. The grain size of this type of data is smaller when compared to the grain size taken from surveys or institutional repositories (Worsley and Blikstein 2013), and represents the velocity component of big data. The data are also voluminous with respect to what is generally gathered in education, such that many measurements of each individual student are taken, providing potentially thousands of data points per student during a given learning activity. This approach has been used to distinguish how experts versus novices approach (and solve) similar tasks. Related work has aggregated data from audio sources; video sources coded for hand and wrist movements; and electro-dermal activation devices (i.e., changes in the skin's resistance to small electrical current due to sweat), in order to distinguish between when students were using "principle based reasoning" versus "example based reasoning" in STEM contexts The latter was associated with better learning outcomes (Worsley and Blikstein 2015).

Big data in learning analytics is not restrained to quantitative data. There is also substantial work that aims to understand how students write. Snow et al. (2015), for example, designed and piloted an Automated Writing Evaluation (AWE) system in order to detect whether or not students could adapt their writing style in accordance with a specific writing task, rather than recycle previously used writing techniques. Since flexibility is a key indicator of strong writing, the AWE was developed to detect such writing *stealthily*, i.e., by embedding the assessments within the software used to capture students' writing in an unobtrusive and automated manner. Results from the study serve as an example of how learning analytics can be used to understand *particular* students, rather than the abstract—and non-existent— "average" students, at scale. These insights from an AWE do not rely on large amounts of human capital, and serve as additional points of information that help teachers act on insights in a targeted (i.e., personalized) way.

When paired with other types of data (e.g., surveys, standardized test results, and demographic information), AWEs can help predict the quality of students' writing. Crossley et al. (2015) used this technique to specify better performing predictive models, and suggested that this technique could be employed to provide more personalized information to students—like an intelligent tutor for writing. Doing so would, hypothetically, result in better feedback that was also faster and personalized for each student (Crossley et al. 2015).

Many of the insights derived through learning analytics techniques are operationalized through the use of "dashboards," which provide instructors with real-time information regarding the learning outcomes of their students (Diana et al. 2017; Verbert et al. 2013; West 2012). Such insights allow stakeholders to detect students who are struggling and to interevene when students are predicted to be at risk of academic challenges or failure (Dawson et al. 2017; Diana et al. 2017; Herodotou et al. 2017; Ocumpaugh et al. 2017).

Learning Analytics and Social Justice

Regardless of the form they take, applications of learning analytics instantiate the goals of various stakeholders in their designs for the purpose of personalizing students' learning experiences. Institutions that implement them embed various decisions regarding what sort of information may or may not help students, as well as the form that information takes. This is no different than the goals of a teacher or instructor; the key difference is speed, scalability, and personalization. Learning analytics-driven educational technologies inform the decision-making process of various stakeholders and extend the decision-making powers of those who design them. This results in the capability to scale information interventions to thousands of students quickly, resulting in dynamic, person-centered learning with the potential to enable more equitable outcomes for students who are at risk of being neglected, thus contributing to social justice in education.

Recent evidence supports this. Brown et al. (2017), for example, detected differences in students who struggled in their coursework. They found that students who moderately struggled in a college course recovered by using tools that helped them organize their studying, whereas tools that helped them prepare for exams better served students who struggled more. In a large university setting, detecting such behaviors is generally not possible. The ability to scale and adapt to thousands of students, while simultaneously providing information to other stakeholders invested in students' learning (Aguilar et al. 2014; Baker and Inventado 2014; Tempelaar et al. 2014; Wise et al. 2011), positions learning analytics as a potential contributor to social justice in education.

With personalization as a starting point, learning analytics has the potential to give students new tools to study themselves as learners and be more reflective of their study habits (e.g., Cutumisu et al. 2015; Roll and Winne 2015), while also connecting them to institutional resources that may have otherwise remainded hidden. Both of these affordances contribute to the potential of learning analytics to contribute to more socially just outcomes for students who might have otherwise been more susceptible to academic failure. Academic advisors, for example, can use a learning analytics-based Early Warning Systems to address students' academic challenges before it becomes too late (e.g., Arnold and Pistilli 2012; Krumm et al. 2014). Such interventions are important for the success of at-risk student populations, especially during their first year (Tinto 1999).

Students' learning experiences, moreover, will drive the algorithms that connect them to the resources they need to learn better and faster. This defacto personalization has the potential to give students more autonomy over aspects of their learning, will teach them to be more reflective of their learning practice, and also will give those invested in a student's learning process more information than was previously available. Thus, learning analytics-driven educational technologies are poised to move educational interventions away from educating the "average" student and towards educating an *actual* student. One can imagine, for example, a learning analytics application that parses a particular students' piece writing, yielding results that speak to his or her flexibility as a writer (Crossley et al. 2015). Armed with this information, the student can practice and receive feedback quickly. The learning analytics application can also suggest a staff person at writing center who the student can contact. It can also suggest other students who might be helpful, thus facilitating collaborative learning. There are numerous possibilities—all of which are an improvement over writing in isolation and waiting for feedback that may take days, or weeks, to arrive. As learning analytics matures, the time between detecting struggle and intervening is decreased (e.g., Fu et al. 2017). This rapid response is necessary for improving the educational outcomes of underserved populations who often struggle in new environments (e.g., Suzuki et al. 2014). By detecting struggling students and giving them a personalized way to understand their own learning or address deficiencies, the likelihood of small differences becoming large ones is decreased.

If students learn in a blended learning environment (Garrison and Kanuka 2004), even more possibilities emerge. As students generate data within digital learning environments, learning analytics techniques can be applied to help students and teachers alike unpack and understand important self-regulated learning processes, such as self-monitoring, which is helpful when it comes to comprehending new material (Zimmerman and Paulsen 1995). This represents a fundamental shift in formal education. Until recently, the feedback students received generally consisted of formative assessments during a course, and summative assessments at the end of a designated leaning period (e.g., an academic term). Such methods, while important, are imperfect solutions that often give students information after it is too late to change ineffective behaviors, or do not capture learning behaviors at a granular enough level. Since learning analytics is predicated on the availability of big data, this no longer needs to be the case. Using the previous example, one can imagine giving young writers the ability to diagnose their own writing habits by providing feedback at a much faster rate. This can be as simple as helping them understand their use of repetitive beginnings in paragraphs. If more data are available, one can also help young writers understand which time of day is their most productive time to write. In the classroom, the insights generated by learning analytics-based applications can help teachers truly differentiate and alleviate some of the burden from new or struggling teachers; both of which have been shown to play a role in underserved student populations (Gusky 2007).

Learning analytics-based applications thus have the potential to help students reflect on their learning while also helping them to connect to resources they might have otherwise not realized were available. In this way learning, standards can remain fixed for every student while also allowing for personalization. Thus, the "finishing line" may be similar across every student (e.g., graduating from high school, understanding order of operations in math, etc.), however the pathways—

the ways in which students achieve the outcomes they set for themselves or are set for them—become individualized. The personalization process is also internal: if students are given feedback tailored to their own goals and experience, then they will have more opportunities to shape their goals and learning practices.

Not a Silver Bullet

Learning analytics-driven educational technologies are not *silver bullets* that can fix education. While learning analytics techniques, practices, and applications can potentially change the way various stakeholders understand student learning, there is still more work to be done. In fact, there are reasons to temper excitement in order to address key challenges faced during the development and implementation of learning analytics-based technologies. Advancements in the collection, analysis, and integration of student data into various applications, for example, come with risks, and also pose ethical dilemmas that need to be understood and addressed as learning analytics-based educational technologies are implemented.

Implicit in their design is the notion that all available data is potentially available for analysis. However, as with other analytic techniques, learning analytics can suffer from a latent variable problem, i.e., an important dimension about student learning may exist, and not be captured. This can occur because the mechanism to capture a certain behavior has not been invented or thought of, or it can be the result of students simply not engaging with the technology designed to capture important behaviors. This challenge does not negate the importance of acting on the information that we *can* capture (Prinsloo and Slade 2017). It also creates an imperative to develop more measures, such as those in development, to better understand the needs of students (Whitelock-Wainwright et al. 2017).

Privacy concerns must also be addressed. A perceived lack of privacy on the part of the students whose data is being used, for example, can undermine the legitimacy of learning analytics applications before they get off the ground (Hoel and Chen 2016). Recent work in data-privacy issues has noted that learning analytics research may rely too heavily on the human-subject research-informed consent paradigm, suggesting that traditional notions of consent cannot account for the continual use of one's data by learning analytics researchers and/or applications (Cormack 2016). When learning analytics is done at a smaller scale, these concerns are compounded because there is no guarantee of privacy to begin with (Rodríguez-Triana et al. 2016). Privacy issues aside, there are also practical dimensions that need to be addressed. When developing learning analytics-based applications, it is important to keep in mind that there will be instances where implementations are associated with unintended consequences (e.g., Lonn et al. 2014). Data visualizations often used in various applications, moreover, are not free of context, and may be laden with racialized content that needs to be negotiated by teachers and students alike (Philip et al. 2016).

There is also the potential to aggregate students' information in ways that are not conceptually dissimilar to an average. Clustering, for example, is a data-driven approach that detects patterns in multiple variables in order to create groups based on similarity (Jain et al. 1999). While more sophisticated than taking an average, clustering still groups individuals together (Slater et al. 2016). The resulting set of groups (or clusters) have the potential to wash out individual behaviors, thus replicating a key issue learning analytics is meant to address.

Conclusion

Returning to Daniels, we see that in the 1950s, designing for a nonexistent average man led to poor outcomes. In 2016 we are still designing educational experiences for the average student, and have room to improve. It is unfair and ineffective to expect teachers to personalize the experiences of each of their students, all at once, without help. With the aid of learning analytics-driven educational technologies, we can move closer to the personalized instruction method championed by Keller (1968)—learning analytics provides a way forward. Utilizing big data, good design, and the input of the stakeholders they are meant to serve, learning analytics techniques aim to develop applications for the sole purpose of reducing the classroom size to 1. This will not happen overnight, or without first addressing ethical and methodological concerns. Over time, however, these digital innovations will enable us to finally do away with a model of education that teaches toward the non-existent average student, replacing it with one that is more socially just and equitable; one that acknowledges and supports the individual needs of every student.

Acknowledgements The author would like to thank the University of Southern California's Provost's Postdoctoral Scholars Program for its support, as well as Dr. Stuart Karabenick, Dr. Steven Lonn, and Dr. Stephanie Teasley who nurtured an early interest in learning analytics applications.

References

Aguilar, S. J., Lonn, S., & Teasley, S. D. (2014). Perceptions and use of an early warning system during a higher education transition program (pp. 113–117). Presented at the LAK '14 Proceedings of the Third International Conference on Learning Analytics and Knowledge, New York, New York, USA: ACM Press. https://doi.org/10.1145/2567574.2567625.

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: learning by doing and explaining with a computer-based cognitive tutor. *Cognitive Science, 26*, 147–179.

Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: a model of help-seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education, 16*, 101–130.

Anderson, J. R., Boyle, C. F., & Reiser, B. J. (1985). Intelligent tutoring systems. *Science, 228*, 456–462.

Arnold, K. E., & Pistilli, M. D. (2012). Course signals at purdue: using learning analytics to increase student success. Presented at the LAK '12: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge.

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

Baker, R. S. J. D., Corbett, A. T., Koedinger, K. R., Evenson, S. E., Roll, I., Wagner, A. Z., et al. (2006). Adapting to when students game an intelligent tutoring system (pp. 392–401). Presented at the Proceedings ITS.

Baker, R. S. J. D., DMello, S. K., Rodrigo, M. M. T., & Grasser, A. (2010). Better to be frustrated than bored: the incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments (pp. 52–63). Presented at the International Conference on User Modeling, Adaptation, and Personalization.

Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research (pp. 309–318). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York: ACM Press. doi: https://doi.org/10.1145/3027385.3027403.

Borland, M. V., Howsen, R. M., & Trawick, M. W. (2005). An investigation of the effect of class size on student academic achievement. *Education Economics, 13*(1), 73–83. https://doi.org/10.1080/0964529042000325216.

Brooks, C., Greer, J., & Gutwin, C. (2014). The data-assisted approach to building intelligent technology-enhanced learning environments. In J. A. Larusson & B. White (Eds.), *Learning analytics: from research to practice* (pp. 123–156). New York: Springer New York. https://doi.org/10.1007/978-1-4614-3305-7_7.

Brown, M. G., DeMonbrun, R. M., & Teasley, S. D. (2017). Don't call it a comeback (pp. 489–493). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York, New York: ACM Press. https://doi.org/10.1145/3027385.3027393.

Chen, M., Mao, S., & Liu, Y. (2014). Big data: a survey. *Mobile Networks and Applications, 19*(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0.

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683–695. https://doi.org/10.1080/13562517.2013.827653.

Cormack, A. (2016). A data protection framework for learning analytics. *Journal of Learning Analytics, 3*(1), 91–106. 10.18608/jla.2016.31.6.

Crossley, S., Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Pssst... textual features... there is more to automatic essay scoring than just you! (pp. 203–207). Presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, New York: ACM Press. https://doi.org/10.1145/2723576.2723595.

Curran, F. C., & Kellogg, A. T. (2016). Understanding science achievement gaps by race/ethnicity and gender in kindergarten and first grade. *Educational Researcher, 45*(5), 273–282.

Cutumisu, M., Blair, K. P., Chin, D. B., & Schwartz, D. L. (2015). Posterlet: a game-based assessment of children's choices to seek feedback and to revise. *Journal of Learning Analytics, 2*(1), 49–71.

Daniels, G. (1952). *The "Average Man"?* (pp. 1–12). Air Force Aerospace Medical Research Lab Wright-Patterson Afb OH.

Darling-Hammond, L. (2000). Teacher quality and student achievement: a review of state policy evidence Linda Darling-Hammond Stanford University. *Education Policy Analysis Archives, 8*(1), 1–44.

Dawson, S., Jovanović, J., Gašević, D., & Pardo, A. (2017). From prediction to impact (pp. 474–478). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York: ACM Press. https://doi.org/10.1145/3027385.3027405.

Dewy, J. (1915). *The school and society.* Chicago: University of Chicago Press.

Diana, N., Eagle, M., Stamper, J., Grover, S., Bienkowski, M., & Basu, S. (2017). An instructor dashboard for real-time analytics in interactive programming assignments (pp. 272–279). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York: ACM Press. https://doi.org/10.1145/3027385.3027441.

Finn, J., & Achilles, C. M. (1990). Answers and questions about class size: a statewide experiment. *American Educational Research Journal, 27*(3), 557–577.

Fu, X., Shimada, A., Ogata, H., Taniguchi, Y., & Suehiro, D. (2017). Real-time learning analytics for C programming language courses (pp. 280–288). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York: ACM Press. https://doi.org/10.1145/3027385.3027407.

Gamoran, A., & Weinstein, M. (1998). Differentiation and opportunity in restructured schools. *American Journal of Education, 106*(3), 385–415.

Garrison, D. R., & Kanuka, H. (2004). Blended learning: uncovering its transformative potential in higher education. *The Internet and Higher Education, 7*(2), 95–105. https://doi.org/10.1016/j.iheduc.2004.02.001.

Gary-Bobo, R. J., & Mahjoub, M.-B. (2013). Estimation of class-size effects, using "Maimonides' Rule" and other instruments: the case of French Junior High Schools. *Annals of Economics and Statistics, 111/112*, 193–225.

Graue, E., Hatch, K., Rao, K., & Oen, D. (2007). The wisdom of class-size reduction. *American Educational Research Journal, 44*(3), 670–700. https://doi.org/10.3102/000283120.

Gusky, T. R. (2007). Closing achievement gaps: revisiting Benjamin S. Bloom's 'Learning for Mastery'. *Journal of Advanced Academics, 19*(1), 8–31.

Herodotou, C., Rienties, B., Boroowa, A., Zdrahal, Z., Hlosta, M., & Naydenova, G. (2017). Implementing predictive learning analytics on a large scale (pp. 267–271). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York: ACM Press. https://doi.org/10.1145/3027385.3027397.

Hoel, T., & Chen, W. (2016). Privacy-driven design of learning analytics applications: exploring the design space of solutions for data sharing and interoperability. *Journal of Learning Analytics, 3*, 139–158. 10.18608/jla.2016.31.9.

IBM. (2017). The Four Vs of Big Data. Retrieved March 29, 2017, from http://www.ibmbigdatahub.com/infographic/four-vs-big-data.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys, 31*(3), 1–60.

Johnson, I. Y. (2010). Class size and student performance at a public research university: a cross-classified model. *Research in Higher Education, 51*(8), 701–723. https://doi.org/10.1007/sll.

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A. (2014). *NMC Horizon Report: 2014 Higher Education Edition* (pp. 1–52). Austin, Texas.

Keller, F. S. (1968). Good-Bye Teacher. *Journal of Applied Behavior Analysis, 1*, 79–89.

Keller, F. S. (1974). Ten years of personalized instruction. *Teaching of Psychology, 1*(1), 1–9.

Krumm, A. E., Waddington, R. J., Teasley, S. D., & Lonn, S. (2014). A learning management system-based early warning system for academic advising in undergraduate engineering. In J. A. Larusson & B. White (Eds.), *Learning analytics: from research to practice* (pp. 103–119). New York: Springer New York. https://doi.org/10.1007/978-1-4614-3305-7_6.

Kulik, J. A., Kulik, C.-L. C., & Cohen, P. A. (1979). A meta-analysis of outcome studies of Keller's personalized system of instruction. *American Psychologist, 34*(4), 307–318.

Ladson-Billings, G. (2006). From the achievement gap to the education understanding achievement in U.S. schools. *Educational Researcher, 35*(7), 3–12.

Lawrence-Brown, D. (2004). Differentiated instruction: inclusive strategies for standards-based learning that benefit the whole class. *American Secondary Education, 32*(3), 34–62.

Lee, J. (2002). Racial and Ethnic achievement gap trends: reversing the progress toward equity? *Educational Researcher, 31*(1), 3–12.

Lee, V. E., & Bryk, A. (1989). A Multilevel Model of the Social Distribution of High School Achievement. *Sociology of Education, 62*(3), 172–192.

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2013). Issues, challenges, and lessons learned when scaling up a learning analytics intervention. Presented at the LAK '13: Proceedings of the Third International Conference on Learning Analytics and Knowledge.

Lonn, S., Aguilar, S. J., & Teasley, S. D. (2014). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior, 47*, 1–8. https://doi.org/10.1016/j.chb.2014.07.013.

McTighe, J., & Brown, J. L. (2005). Differentiated instruction and educational standards: is detente possible? *Theory Into Practice, 44*(3), 234–244. https://doi.org/10.1207/s15430421tip4403_8.

Miller, W. L., Baker, R. S., Labrum, M. J., Petsche, K., Liu, Y.-H., & Wagner, A. Z. (2015). Automated detection of proactive remediation by teachers in reasoning mind classrooms (pp. 290–294). Presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, New York: ACM Press. https://doi.org/10.1145/2723576.2723607.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: a classroom study of values affirmation. *Science, 330*(6008), 1234–1237. https://doi.org/10.1126/science.1195996.

Murray, T. (1999). Authoring intelligent tutoring systems: an analysis of the state of the art. *International Journal of Artificial Intelligence in Education, 10*, 98–129.

Ocumpaugh, J., Baker, R. S., San Pedro, M. O. C. Z., Hawn, M. A., Heffernan, C., Heffernan, N., & Slater, S. A. (2017). Guidance counselor reports of the assessments college prediction model (ACPM) (pp. 479–488). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York: ACM Press. https://doi.org/10.1145/3027385.3027435.

Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily online testing in large classes: boosting college performance while reducing achievement gaps. *PLoS One, 8*(11), e79774–e79776. https://doi.org/10.1371/journal.pone.0079774.

Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction, 34*(4), 361–388. https://doi.org/10.1080/07370008.2016.1210418.

Polikoff, M. S. (2012). Instructional Alignment under No Child Left Behind. *American Journal of Education, 118*(3), 341–368. https://doi.org/10.1086/664773.

Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room (pp. 46–55). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York: ACM Press. https://doi.org/10.1145/3027385.3027406.

Quinn, D. M. (2015). Kindergarten black-white test score gaps: re-examining the roles of socioeconomic status and school quality with new data. *Sociology of Education, 88*(2), 120–139. https://doi.org/10.1177/0038040715573027.

Raghupathi, W., & Raghupathi, V. (2014). Big data analytics in healthcare: promise and potential. *Health Information Science and Systems, 2*(1), 211–210. https://doi.org/10.1186/2047-2501-2-3.

Rodríguez-Triana, M. J., Martínez-Monés, A., & Villagrá-Sobrino, S. (2016). Learning analytics in small-scale teacher-led innovations: ethical and data privacy issues. *Journal of Learning Analytics, 3*(1), 43–65. 10.18608/jla.2016.31.4.

Roll, I., & Winne, P. H. (2015). Understanding, evaluating, and supporting self-regulated learning using learning analytics. *Journal of Learning Analytics, 2*(1), 7–12.

Scales, P. C., Roehlkepartain, E. C., Neal, M., Kielsmeier, J. C., & Benson, P. L. (2006). Reducing academic achievement gaps: the role of community service and service-learning. *The Journal of Experimental Education, 29*(1), 38–60.

Shin, Y. (2012). Do black children benefit more from small classes? Multivariate instrumental variable estimators with ignorable missing data. *Journal of Educational and Behavioral Statistics, 37*(4), 543–574. https://doi.org/10.3102/1076998611431083.

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

Slater, S., Joksimovic, S., Kovanovic, V., Baker, R. S., & Gasevic, D. (2016). Tools for educational data mining: a review. *Journal of Educational and Behavioral Statistics*, 1–22. https://doi.org/10.3102/1076998616666808.

Snow, E. L., Allen, L. K., Jacovina, M. E., Perret, C. A., & McNamara, D. S. (2015). You've got style (pp. 194–202). Presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, New York: ACM Press. https://doi.org/10.1145/2723576.2723592.

Stiggins, R., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory Into Practice, 44*(1), 11–18. https://doi.org/10.1207/s15430421tip4401_3.

Suthers, D., & Rosen, D. (2011). A unified framework for multi-level analysis of distributed learning (pp. 1–11). Presented at the LAK '11: Proceedings of the 1st International Conference on Learning Analytics and Knowledge.

Suzuki, A., Amrein-Beardsley, A., & Perry, N. J. (2014). A summer bridge program for underprepared first-year students: confidence, community, and re-enrollment. *Journal of the First-Year Experience Students in Transition, 2*(24), 85–106.

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2014). In search for the most informative data for feedback generation: learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 1–11. https://doi.org/10.1016/j.chb.2014.05.038.

Teplovs, C., Fujita, N., & Vatrapu, R. (2011). Generating predictive models of learner community dynamics (pp. 1–6). Presented at the LAK '11: Proceedings of the 1st International Conference on Learning Analytics and Knowledge.

Tinto, V. (1999). Taking retention seriously: Rethinking the first year of college. *NACADA journal, 19*(2), 5–9.

Trkman, P., McCormack, K., de Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems, 49*(3), 318–327. https://doi.org/10.1016/j.dss.2010.03.007.

Vanneman, A., Hemphill, F. C., & Rahman, T. (2011). Achievement gaps: how hispanic and white students in public schools perform in mathematics and reading on the national assessment of educational progress. National Center for Education Statistics (Vol. NCES, pp. 2011–2459).

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*(10), 1500–1509. https://doi.org/10.1177/0002764213479363.

West, D. M. (2012). *Big data for education: data mining, data analytics, and web dashboards*. Brookings (pp. 1–11).

Whitelock-Wainwright, A., Gašević, D., & Tejeiro, R. (2017). What do students want? (pp. 368–372). Presented at the Seventh International Learning Analytics & Knowledge Conference, New York: ACM Press. https://doi.org/10.1145/3027385.3027419.

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics (pp. 203–211). Presented at the LAK '14 Proceedings of the Third International Conference on Learning Analytics and Knowledge, New York: ACM Press. https://doi.org/10.1145/2567574.2567588.

Wise, A. F., Saghafian, M., & Padmanabhan, P. (2011). Towards more precise design guidance: specifying and testing the functions of assigned student roles in online discussions. *Educational Technology Research and Development, 60*(1), 55–82. https://doi.org/10.1007/s11423-011-9212-7.

Wise, A. F., Zhao, Y., & Hausknecht, S. N. (2014). Learning analytics for online discussions: embedded and extracted approaches. *Journal of Learning Analytics, 1*, 48–71.

Worsley, M., & Blikstein, P. (2013). Towards the development of multimodal action based assessment. Presented at the LAK '13: Proceedings of the Third International Conference on Learning Analytics and Knowledge.

Worsley, M., & Blikstein, P. (2015). Leveraging multimodal learning analytics to differentiate student learning strategies (pp. 360–367). Presented at the Proceedings of the Fifth International Conference on Learning Analytics And Knowledge, New York: ACM Press. https://doi.org/10.1145/2723576.2723624.

Zimmerman, B. J., & Paulsen, A. S. (1995). Self-monitoring during collegiate studying: an invaluable tool for academic self-regulation. *New Directions for Teaching and Learning, 63*, 13–27.