Learning, Schooling, and Data Analytics

Ryan S. J. d. Baker

Baker, R. J. d. (2013). Learning, schooling, and data analytics. In M. Murphy, S. Redding, & J. Twyman (Eds.), *Handbook on innovations in learning* (pp. 1χω–19π). Philadelphia, PA: Center on Innovations in Learning, Temple University; Charlotte, NC: Information Age Publishing. Retrieved from http://www.centeril.org/

Learning, Schooling, and Data Analytics

Ryan S. J. d. Baker

Since the 1960s, methods for extracting useful information from large data sets, termed *analytics* or *data mining*, have played a key role in fields such as physics and biology. In the last few years, the same trend has emerged in educational research and practice, an area termed *learning analytics* (LA; Ferguson, 2012) or *educational data mining* (EDM; Baker & Yacef, 2009). In brief, these two research areas seek to find ways to make beneficial use of the increasing amounts of data available about learners in order to better understand the processes of learning and the social and motivational factors surrounding learning. The goal of these efforts is to produce more efficient, more effective, and deeper learning in the context of increasingly positive learning experiences.

The emergence of EDM/LA is a recent phenomenon. The first meetings of scientists in this area were the Educational Data Mining workshops, which started in 2005 and became an annual conference series in 2008. This conference series was joined by the Learning Analytics and Knowledge conference series in 2011. The two research areas of EDM and LA, emerging from different communities of scientists and practitioners, have somewhat different goals; discussing these differences is outside the scope of this report (see Siemens & Baker, 2012). In brief, the validity of models of learners and learning is perhaps the key focus of the EDM community, whereas the use of the results of analysis to drive changes in practice by instructors is perhaps the key focus of the LA community. The conferences in EDM and LA were followed by the establishment of journals devoted to the topics, with the *Journal of Educational Data Mining* commencing publication in 2009 and the *International Journal of the Society for Learning Analytics Research* expected to commence publication in 2013. As of this writing, the

International Educational Data Mining Society has approximately 150 members and over 600 subscribers on its mailing lists.

A range of methods has been developed by these two communities, drawing from areas such as data mining, computational science, statistics, psychometrics, and social network analysis. (A selection of these methods will be discussed below; a fuller review can be found in Baker & Siemens, in press).

Research Synthesis

The methods of EDM have been applied to accomplish a range of objectives. This section reviews some of the applications which have had relatively large impacts or have relatively large potential, focusing on applications of particularly strong relevance to the readers of this *Handbook*.

One of the first applications of EDM was the development of models that could infer a student's knowledge as he or she worked through educational software. These inferences are in turn used to drive adaptation by the system. This application in fact preceded the existence of EDM or LA as research areas. Though student knowledge modeling began as a research area in the 1970s (Goldstein, 1979), the first model, which was both based on automated exploration of data and which achieved widespread dissemination in educational software, was Corbett and Anderson's (1995) Bayesian knowledge tracing (BKT) algorithm. One key difference between this algorithm and the types of student knowledge modeling used previously by the psychometrics community, for example in testing, was that BKT explicitly accounts for the fact that the student is learning at the same time he or she is being assessed; in other words, student knowledge is treated as a moving target. BKT was then incorporated into Cognitive Tutor software curricula for algebra and geometry (Koedinger & Corbett, 2006), sold by Carnegie Learning Inc., which was used by around 5% of U.S. high school students each year throughout the first decade of the 2000s. This software used BKT to decide when to advance the student on to new material, implementing an approach termed "mastery learning" (Bloom, 1968), in which the student does not advance until he or she demonstrates proficiency. By integrating BKT into mastery learning, Cognitive Tutor Algebra I was able to improve student test scores, with replication, in a range of settings (Koedinger & Corbett, 2006), although performance for geometry has been more mixed (Pane, McCaffrey, Slaughter, Steele, & Ikemoto, 2010). It is important to note that the automated algorithms and learning support in Cognitive Tutor replaced the workbook rather than the teacher; in Cognitive Tutor classrooms, the teacher spends more time interacting with students in one-on-one learning support sessions than in full-class teaching (Schofield, 1995), perhaps another reason for this approach's success.

Since the implementation of Cognitive Tutors, new online learning systems have added emphasis on providing actionable and formative information to teachers. For example, the ASSISTment system (it "assesses while it assists") has created a reporting system that teachers can use to determine both what material specific students are struggling with and what items the entire class is struggling with (Feng & Heffernan, 2006). Teachers using the system review student homework before class and are able to change the focus of classroom activities based on data on student understanding, leading to better classroom performance than is seen with traditional homework (Koedinger, McLaughlin, & Heffernan, 2010; Mendicino, Razzaq, & Heffernan, 2009).

The types of formative information that can be assessed by online learning systems have gone beyond just student knowledge in recent years. Algorithms for assessing disengaged behaviors have been developed for learning systems recently (Baker, 2007; Baker, Corbett, & Koedinger, 2004; Pardos, Baker, San Pedro, Gowda, & Gowda, 2013; San Pedro, Baker, & Rodrigo, 2011), making it possible to assess with reasonable accuracy whether students are careless, off-task, or intentionally misusing educational software, among other disengaged behaviors. These algorithms have been extended to also infer student emotion during learning, just from data readily available to computer systems (i.e., no physiological sensors; see Baker et al., 2012; Pardos et al., 2013; Sabourin, Rowe, Mott, & Lester, 2011). As these models are built into systems such as ASSISTments or Crystal Island, increasing amounts of information will be available to classroom teachers; the key challenge will be providing it to teachers in useful and timely fashions.

> ***Teachers using the system review student homework before class and are able to change the focus of classroom activities based on student understanding...***

Another direction for integrating EDM and LA research into educational practice is to predict student dropout and course failure, a step towards providing early intervention. One particularly successful example is the Purdue Signals Project, reported to have significantly improved student outcomes at Purdue University (Arnold, 2010). This system uses prediction models to infer early in the semester which students are likely to fail or drop out of a course; a list of students at risk is generated and sent to an instructor, along with recommended template emails for these students which inform them about help resources available. This type of system is being implemented at an increasing number of universities, both in independent projects (Ming & Ming, 2012) and through a commercial vendor, Ellucian, which is distributing the Signals software to additional universities.

While dropout and failure prediction at the K–12 level have not yet reached the level of deployment and demonstrated success of the Purdue Signals Project, there are several examples of successful prediction of student dropout at the K–12 level. To give just a few examples, Tobin and Sugai (1999) predict high school dropout from middle school disciplinary records; Bowers (2010) uses changes in student achievement to predict high school dropout as early as third grade; San Pedro, Baker, Bowers, and Heffernan (in press) use data on middle school student emotion and learning within the aforementioned ASSISTment system to predict which students will attend college. Each of these approaches has the potential to be used at scale; the challenges to doing so are organizational rather than technical.

Beyond supporting specific changes in practice, EDM and LA research has played an increasingly important role in supporting basic discovery in education research. The opportunity to leverage very fine-grained data (often multiple data points per minute) across entire years of data for a specific student, in combination with automated methods for sifting through that data, has been an excellent opportunity for better understanding learners and learning. Types of EDM methods, such as discovery with models and structure discovery algorithms, have enabled a variety of analyses, including discovery of which exploratory learning strategies are most effective (Amershi & Conati, 2009), which patterns of group work lead to more successful group projects (Perera, Kay, Koprinska, Yacef, & Zaiane, 2009), which meta-cognitive behaviors lead to deep learning (Baker et al., 2012), and how small-scale choices in the design of educational software can lead to substantial differences in student engagement (Baker et al., 2009).

> ***The opportunity to leverage very fine-grained data...across entire years of data for a specific student...has been an excellent opportunity for better understanding learners and learning.***

Action Principles

In this section, I propose a set of action principles for schools, local education agencies (LEAs), and state education agencies (SEAs), suggesting how the emerging fields of learning analytics and educational data mining can be used to improve their practice.

Action Principles for Schools

Provide formative data to teachers on student learning. In recent years, the advent of learning systems such as ASSISTments (but also Cognitive Tutors, Reasoning Mind, Aleks, LearnBop, and many others) has presented an opportunity to provide teachers with considerably more information on their students' learning, generally in easy-to-interpret formats. Depending on a school's goals, some of these systems (such as Cognitive Tutors and Reasoning Mind) can be adopted as an entire curriculum; others, such as ASSISTments and LearnBop, simply replace existing homework or seatwork and can be used with a variety of curricula.

These systems provide teachers with information on which students are struggling and what they are struggling on. This enables teachers to identify what material these students need support with, so that the teacher can provide them with extra assistance (Schofield, 1995). Sometimes, a teacher can also see by using these systems that a specific topic is difficult for all students; this is also possible to determine when the teacher grades by hand, but the teacher is informed earlier by automated systems, thus supporting timely intervention.

Predict which students are at risk for dropping out. As discussed above, one of the key successes of learning analytics at the undergraduate level is predicting which students are at risk of failing or dropping out. At that level, success has been achieved not only in predicting who is at risk but also in embedding this information in effective interventions used to reduce dropout (Arnold, 2010).

Several research projects have demonstrated that the same type of prediction is possible for K–12 schools. The work by Bowers (2010) in predicting high school dropout from grades students receive in elementary school demonstrates that this type of prediction is possible just from the data already available in schools. Similarly, data on disciplinary referrals (e.g., fighting) during middle school can predict who will drop out in high school (Tobin & Sugai, 1999). However, both of these types of indicators may be identifying students at very high risk, students whose problems are outside those that are easily addressed by schools. Dropout prediction from interactions with educational software may provide a way to identify at-risk students whose challenges can be more easily addressed, and may provide more precise information on the factors causing those students to be at risk. For example, recent work has indicated that educational software can infer not just student knowledge but also multiple dimensions of student engagement. Long-term prediction from educational software is still emerging (San Pedro et al., in press) but is likely to be available in an increasing number of educational software packages used in schools in the years to come.

Identify learning topics that are being learned less well within school. Recent educational software is able to identify which skills and topics are being learned less well than others within a specific classroom or school. This type of information is available in reports from many modern learning software packages, including but not limited to ASSISTments, the Cognitive Tutor, LearnBop, and Reasoning Mind. This type of information does not require using a software package—it is possible to think of teachers across schools recording homework data, tagging it by topic, and looking together for topics where performance is poor—but it is much easier to do in schools and classrooms that use educational software since the bookkeeping and data integration is offloaded to a computer system.

Understanding the topics for which a school's current curriculum and pedagogical approaches are working less effectively creates opportunities to redesign teaching in those areas or to supplement current practice with other resources. If a school is generally performing poorly on division of fractions across teachers, for example, it is probably not a flaw in one person's teaching but instead a flaw in the curriculum being used, a flaw that can be addressed throughout the school.

Capture and respond to changes in student engagement. In 2013, the automated assessment of student engagement and emotion remains primarily within research classrooms, but it is emerging within a range of learning systems, making it likely that it will become generally available in classrooms in the coming years. As automated assessment of student engagement and emotion becomes increasingly feasible to integrate within online learning systems, such as ASSISTments, it is likely to become useful to teachers. When it is available, teachers and school psychologists will be able to use it to identify early students who have become disengaged across classes, potentially identifying a student in need of an intervention. If problem behaviors are below the threshold of office referrals, a student's general changes in behavior may not be noticed; with this type of technology, it may be possible to identify shifts in engagement quickly. Even within a single class, emotion- and engagement-sensing technology may prove quite useful. For example, if a teacher can identify that a student was frustrated during his or her online homework the night before, it may be possible to talk to the student to better understand why the material was particularly difficult.

Action Principles for Local Education Agencies

Identify specific areas of excellence and high success in teaching practice. When educational software that assesses engagement and learning is used in schools, it can be beneficial not just to individual teachers and schools but to local education agencies (LEAs) as well. This type of assessment can provide information that can help LEAs to identify teachers that are successful in promoting engagement and learning in specific areas. The expertise these teachers have can then be leveraged by their LEA. For example, if a teacher is succeeding at teaching a topic that other teachers are known to struggle with—as manifested by better performance by his/her students on that topic—that teacher could give a brief workshop on his or her teaching strategies. Similarly, if one teacher's classes generally experience less boredom (while learning equally well), it may be worth having this teacher mentor other teachers in engaging their students. In this fashion, it may be possible to identify exact areas of excellence and share them across a school district.

Although these methods can be used to identify exemplary teachers, rewarding teachers who are particularly successful according to these types of internal measures may have undesired effects. If teacher pay were linked to evidence of frustration in a system like ASSISTments, some teachers might alter their classroom practice in undesirable ways, to try to "game the system," for example, by walking around the classroom, immediately giving answers to every struggling student. Even if this did not improve the assessment of engagement by the software, it might still be attempted, with unpredictable and likely undesirable results. Automated detectors will be more effective, and more useful, if there is not an attempt to subvert them (necessitating automated detectors of subversion, as seen in Baker et al., 2004). In sum, integrating automated assessment systems into reward structures has the potential to reduce their effectiveness for other goals.

Identify students who could benefit from enrichment programs. Another upcoming opportunity for LEAs is to identify specific students who could benefit from enrichment programs. Across the U.S., after-school, weekend, and summer programs are available to learners, funded by federal agencies—such as the National Science Foundation's Innovative Technology Experiences for Students and Teachers (ITEST) program—state agencies, foundations, and private funders. However, there remains insufficient capacity to provide enrichment programs to all students who want to enroll in them, and the students who do enroll are often drawn from wealthier groups (Gardner, Roth, & Brooks-Gunn, 2009). In addition, not all enrichment programs are the same; there is a question of fit when selecting students for an enrichment program.

When technology becomes readily available to assess engagement in class, it will be increasingly possible to identify students who are highly engaged in specific subjects. These students—especially if they are disadvantaged—should be particularly strong candidates for enrichment programs, and efforts should be made to place them in enrichment programs that fit their interests and will help them develop their interests in these specific areas.

Develop internal expertise in learning analytics. A third recommendation for local education agencies is to develop internal expertise in learning analytics. In recent years, there has been an explosion of data that can be used for a wide range of purposes, as indicated in the recommendations above (both those for schools and for local education agencies). Local education agencies can play an essential role in fulfilling both of these recommendations, conducting analyses at the district level and supporting schools in conducting school-level analyses (or even conducting analyses for schools).

The cost of hiring one or more learning analytics experts or of training an existing member of the LEA in learning analytics methods may in the future be seen as a relatively small expense in relation to the benefits that can be achieved. There are increasing opportunities to train LEA personnel, including the upcoming fall 2013 massive online open course (MOOC) within Coursera, Big Data in Education, and an annual MOOC on learning analytics provided by the Society for Learning Analytics Research. Also, an increasing number of graduate programs specialize in this area. As of this writing, programs in learning analytics or related areas are offered at Teachers College Columbia University, Carnegie Mellon University, and Worcester Polytechnic Institute, programs creating an increasing pool of trained individuals who can provide this type of expertise to schools.

Develop data management and sharing plans to support partnerships with university researchers in line with legal obligations. Beyond hiring their own staff in learning analytics, school districts may be able to leverage the expertise of universities. There is growing pool of university faculty, postdoctoral researchers, and graduate students who are deeply interested in learning analytics and EDM and want to use these methods to benefit American education, at a wide range of institutions, even beyond those officially offering training in these areas. These researchers are a resource that LEAs can leverage to conduct analyses beyond their own capacities. Such collaborations are likely to benefit all but the largest and wealthiest school districts; even for those districts, there may be expertise in learning analytics located in specific university research groups that is duplicated nowhere else.

However, these collaborations will not occur unless appropriate institutional, legal, and infrastructural arrangements are made. One key step is the creation of procedures for quickly de-identifying data sets (removing all potentially identifying information) so they can be shared with university researchers without violation of relevant federal privacy laws and guidelines, such as the Family Educational Rights and Privacy Act (FERPA), the federal law that protects the privacy of student education records. Creating procedures for sending de-identified data to researchers but being able to link findings from those researchers back to individual students within the LEA will be essential in order to benefit those students, using the information obtained in research. Policies for such de-identification would prevent identifiable information from being transmitted outside the school district and designate some individual within the LEA to hold a strictly guarded key, so that the findings can be tracked back to students within the LEA. In addition, LEAs should instruct their institutional review boards to follow relevant federal law and guidelines for fast-tracking research with minimal risk of harm to students, for research projects classified as exempt from review or fit for expedited review under the federal guidelines. Currently, many LEAs—particularly in larger cities—choose not to follow federal guidelines for review of research, instead creating onerous review processes that lead many research groups to avoid working with those LEAs. The result is that students in suburban school districts benefit more from the university researchers in major urban centers than students in those urban centers, reinforcing inequities. Even after approving research, many LEAs currently require extensive legal agreements, again well beyond federal or state requirements, delaying or preventing research collaborations. In general, streamlining procedures for learning analytics research (while following all federal laws and guidelines, and protecting student privacy) is likely to benefit students considerably and facilitate the task of LEAs in supporting their students.

Action Principles for State Education Agencies

Capture data about students according to broad-based range of indicators. One important way that state education agencies (SEAs) can support learning analytics is by taking steps to collect a broad range of types of data. Many types of data are now available about learners and schools beyond what routinely make it to state education agencies—from log files, to automated assessments, to data from classroom observations. By having a range of types of data, SEAs will be able to conduct analyses of the factors leading to better performance on state standardized exams, higher college attendance, and so on. States should partner with resource centers to select which indicators to capture and encourage vendors to provide understandable and reasonably complete data to their SEA as a condition. Similarly, SEAs should incentivize schools and LEAs to also collect a broad range of data and provide it to SEAs. An SEA's data is unlikely to reach its full potential except in partnership with LEAs that are collecting a broad range of useful data.

While different schools may collect and use data that is not fully compatible, making sure that all of this data is available at the state level will be a useful step towards supporting state-level analyses. For example, even if one learning system tends to predict higher engagement than another learning system, having data from both learning systems will make it possible to see statewide trends.

Form practices for aligning student data even in the face of mobility. School mobility is a fact of 21st-century education; because American society is highly mobile, students are likely to change schools repeatedly during their education. While school mobility may not be problematic for students of high socioeconomic status (SES), it is associated with poorer outcomes among lower SES and minority students, especially if a student changes schools several times (Xu, Hannaway, & D'Souza, 2009). Mobility can also be a problem for tracking students and applying learning analytics to the data from these students; it is easier to obtain data for and therefore apply predictive models to students who do not change school districts, implying that prediction of at-risk status will be least effective for students who are already at risk due to their mobility.

State education agencies can play a key role in tracking these students by using state-level identifiers to track student progress even if the student moves. Equally importantly, SEAs should encourage LEAs and schools to store all data in terms of state-level identifiers and should support LEAs and schools in obtaining student data from other LEAs and schools (ideally through state-level databases that all LEAs provide data to and draw from). In that fashion, learning analytics analyses can leverage all of the data available for a specific student.

States can further support local districts by identifying effective practices for forming partnerships with university and corporate researchers focused on data use. As discussed above, several benefits may accrue to LEAs in forming partnerships with university researchers. SEAs have a key role to play in setting the tone for collaboration and nudging LEAs to develop and conduct these partnerships appropriately. SEAs should educate LEAs about—and encourage them to follow—federal and state guidelines so that LEAs avoid unnecessary and unproductive roadblocks which prevent interventions that would benefit students, while also avoiding violating federal or state laws or violating student privacy.

Identify exemplary teachers and schools. SEAs, even more than LEAs, have the potential to identify schools or teachers who are succeeding at promoting engagement and learning in specific areas. Across a state, there are likely to be exemplary practices, often in unexpected places, that can be identified through learning analytics. These practices can then be studied and communicated across the state in collaboration with resource centers. It is worth noting that, as with LEAs, the indicators that are useful for these types of analyses are better used in a formative fashion than to drive financial incentives (or firings); the incentives for gaming the system or even cheating are substantial if financial incentives are used and doing so would reduce the potential for disseminating exemplary practices statewide.

Identify regional gaps in enrichment programs. As discussed above, enrichment programs are not currently available to all students who want to enroll in them, and the students who do enroll are often drawn from wealthier groups (Gardner, Roth, & Brooks-Gunn, 2009), in part due to regional disparities. While some of the factors leading to these differences are difficult to address (e.g., parental choice and funding choices made by private foundations and individuals), better data on where the needs are may help to influence the allocation of government resources and potential private funding as well. By identifying the number of at-risk students and students likely to benefit from programs, and comparing these numbers to the availability of program slots in different regions, SEAs will be able to identify which regions have an insufficient quantity of enrichment programs and support program expansion and creation. Simply publishing data on where needs exist is likely to influence funding decisions, not just by private foundations and individuals but by programs funded by the federal government. For example, federal programs like National Science Foundation's ITEST might be more likely to fund programs in specific regions declared in need by SEAs than in regions shown to have a relative oversupply of enrichment programs.

Learning analytics may also have the potential to identify more quickly which enrichment programs are working. If an enrichment program is provided to elementary school students, any evidence of its effect on high school dropout rates or college attendance is a distant prospect. Obtaining data on learning and engagement from schools in the year following a student's participation in an enrichment program may provide more rapid signals as to which programs are succeeding in their goals.

Acknowledgements

The author would like to thank Lisa Rossi, Janet Twyman, Marilyn Murphy, and Stephen Page for helpful comments and suggestions.

References

Amershi, S., & Conati, C. (2009). Combining unsupervised and supervised machine learning to build user models for exploratory learning environments. *Journal of Educational Data Mining, 1*(1), 71–81.

Arnold, K. E. (2010). Signals: Applying academic analytics. *Educause Quarterly, 33*, 1–10.

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

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). *Detecting student misuse of intelligent tutoring systems*. Proceedings of the 7th International Conference on Intelligent Tutoring Systems, 531–540.

Baker, R. S. J. d., de Carvalho, A. M. J. A., Raspat, J., Aleven, V., Corbett, A. T., & Koedinger, K. R. (2009). *Educational software features that encourage and discourage "gaming the system."* Proceedings of the 14th International Conference on Artificial Intelligence in Education, 475–482.

Baker, R. S. J. d., Gowda, S. M., Wixon, M., Kalka, J., Wagner, A. Z., Salvi, A.,…Rossi, L. (2012). *Sensor-free automated detection of affect in a Cognitive Tutor for algebra*. Proceedings of the 5th International Conference on Educational Data Mining, 126–133.

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

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

Bloom, B. S. (1968). Learning for mastery. *Evaluation Comment, 1*(2), 1–12.

Bowers, A. J. (2010). Analyzing the longitudinal K–12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out, and hierarchical cluster analysis. *Practical Assessment, Research & Evaluation (PARE), 15*(7), 1–18.

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

Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the ASSISTment system. *Technology, Instruction, Cognition, and Learning Journal, 3*(1–2), 1–8.

Ferguson, R. (2012). *The state of learning analytics in 2012: A review and future challenges* (Tech. Rep. No. KMI-12-01). Milton Keynes, UK: Open University, Knowledge Media Institute. Retrieved from http://kmi.open.ac.uk/publications/techreport/kmi-12-01

Gardner, M., Roth, J. L., & Brooks-Gunn, J. (2009, October). *Can after-school programs help level the playing field for disadvantaged youth?* (Equity Matters: Research Review, 4). New York, NY: Campaign for Educational Equity, Teachers College, Columbia University.

Goldstein, I. J. (1979). The genetic graph: a representation for the evolution of procedural knowledge. *International Journal of Man-Machine Studies, 11*(1), 51–77.

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). New York: Cambridge University Press.

Koedinger, K. R., McLaughlin, E. A., & Heffernan, N. T. (2010). A quasi-experimental evaluation of an on-line formative assessment and tutoring system. *Journal of Educational Computing Research, 43*(4), 489–510.

Mendicino, M., Razzaq, L., & Heffernan, N. T. (2009). A comparison of traditional homework to computer-supported homework. *Journal of Research on Technology in Education, 41*, 331–358.

Ming, N. C., & Ming, V. L. (2012). *Predicting student outcomes from unstructured data*. Proceedings of the 2nd International Workshop on Personalization Approaches in Learning Environments, 11–16.

Pane, J. F., McCaffrey, D. F., Slaughter, M. E., Steele, J. L., & Ikemoto, G. S. (2010). An experiment to evaluate the efficacy of Cognitive Tutor Geometry. *Journal of Research on Educational Effectiveness, 3*(3), 254–281.

Pardos, Z. A., Baker, R. S. J. d., San Pedro, M. O. C. Z., Gowda, S. M., & Gowda, S. M. (2013). *Affective states and state tests: Investigating how affect throughout the school year predicts end of year learning outcomes*. To appear in Proceedings of the 3rd International Conference on Learning Analytics and Knowledge.

Perera, D., Kay, J., Koprinska, I., Yacef, K., & Zaiane, O. R. (2009). Clustering and sequential pattern mining of online collaborative learning data. *IEEE Transactions on Knowledge and Data Engineering, 21*(6), 759–772.

Sabourin, J., Rowe, J., Mott, B., & Lester, J. (2011). *When off-task in on-task: The affective role of off-task behavior in narrative-centered learning environments*. Proceedings of the 15th International Conference on Artificial Intelligence in Education, 534–536.

San Pedro, M. O. C., Baker, R., & Rodrigo, M. M. (2011). *Detecting carelessness through contextual estimation of slip probabilities among students using an intelligent tutor for mathematics*. Proceedings of 15th International Conference on Artificial Intelligence in Education, 304–311.

San Pedro, M. O. Z., Baker, R. S. J. d., Bowers, A. J., & Heffernan, N. T. (in press). *Predicting college enrollment from student interaction with an intelligent tutoring system in middle school*. To appear in Proceedings of the 6th International Conference on Educataional Data Mining.

Schofield, J. W. (1995). *Computers and classroom culture*. Cambridge, MA: Cambridge University Press.

Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254.

Tobin, T. J., & Sugai, G. M. (1999). Using sixth-grade school records to predict violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders, 7*(1), 40–53.

Xu, Z., Hannaway, J., & D'Souza, S. (2009). *Student transience in North Carolina: The effect of school mobility on student outcomes using longitudinal data* (CALDER Working Paper 22). Washington, DC: The Urban Institute.