IJPCE
International Journal of Physics and Chemistry Education, 11(1), 13-17

www.ijpce.org
ISSN: 2589-8876

It's All in the Data - But What is It? Learning Analytics and Data Mining of Multimedia Physics Courses

Gerd Kortemeyer
Michigan State University, United States
kortemey@msu.edu

Abstract
Before "Learning Analytics" became the buzzword that it is today, fueled by the advent of "big data" MOOCs and increased institutional attention to retention and time-to-degree, fine-grained transactions within courses were analyzed as a unique window into student learning: what resources do learners access in which order, how much time do they spend with each resource, to what degree do different resources contribute to success in formative assessment, what is the effect of online discussion forums, how many attempts does it take which learners to solve online homework, and what is the extend and effect of unproductive behavior such as procrastination, guessing, and copying? We will focus on virtual and blended physics courses and use data mining methods to extract various predictive signatures of eventual learner success on exams. We also use classical test theory and item response theory to estimate both quality parameters of online homework and latent learner traits such as ability and the propensity to copy or guess answers. Based on these findings, we will make recommendations for course design.

Keywords: Learning analytics, Data mining, Online, Hybrid

◆Received 13 November 2017 ◆ Revised 30 December 2017 ◆ Accepted 20 November 2018

Introduction
Massive Open Online Courses (MOOCs) with their "Big Data" have recently brought renewed attention to course-level Learning Analytics (e.g., Breslow et al., 2013; Clow, 2013; Kizilcec, 2013), while the general trend had been toward institutional-level Learning Analytics before (e.g., Campbell et al., 2007). However, even long before the advent of MOOCs, course management systems (or any other web-based teaching platforms) have been collecting large amounts of transactional and log data on student interaction and transactions useful for Learning Analytics (e.g., Zaïane, 2002; Bruckmann, 2006; Peñalvo et al.,2011; Psaromiligkos et al., 2011; Mazza et al., 2012). Regular university physics courses make use of online components in a variety of ways: fully online, for-credit "virtual" physics courses have been offered in addition to blended, hybrid, and flipped courses for decades (e.g., Kortemeyer, 2014b); even otherwise traditional courses may have their textbook materials in an online format. Course-level Learning Analytics can be applied to all of these course delivery modes.

Learning Analytics arguably started out as Educational Data Mining (EDM) in the mid-nineties (Romero & Ventura, 2007; Romero & Ventura, 2013), even though EDM is technically different from Learning Analytics and Knowledge (LAK) (Siemens & Baker, 2012). There are subtle differences between EDM and LAK: a stronger focus on automation in EDM versus a stronger focus on informing educators and human judgment in LAK, a reductionist approach in EDM versus a holistic approach in LAK, and frequently different sets of algorithms (Papamitsiou & Economides, 2014; Chatti et al., 2014). – however, the goal is the same: increased learning outcomes (Greller & Drachsler, 2012; Gašević et al., 2015).

Course-level learning analytics put a strong emphasis on formative assessment, i.e., the assessment that accompanies the ongoing learning process. Physics and other STEM disciplines have traditionally been at the forefront of deploying Learning Analytics and Educational Data Mining (Fournier et al., 2011); in fact, physicists may have "discovered" Learning Analytics before the learning sciences did (Baker & Inventado, 2014), likely because STEM disciplines such as physics already traditionally deal with large masses of data, and some of those mechanisms can be brought to bear on educational problems (Mislevy et al., 2012).

Early on, the goal often was the creation of adaptive, personalized, "intelligent" learning environments; this turned out to be an uphill battle, as teaching and learning have many "moving parts" (Gašević et al., 2016). Instead of automated systems that attempt to guide learners independent of instructors (e.g., Khribi, et al., 2008), more frequently dashboards are providing instructors with Learning Analytics (Duval, 2011) as integrated or pluggable components of course management systems (Verbert et al., 2013). A major goal here is the early identification of students-at-risk (Mattingly et al., 2012; Wolff et al., 2013), as well as quality control of online materials (e.g., Dyckhoff et al., 2012) and Just-In-Time Teaching interventions (Novak, 1999). In addition, course-level Learning Analytics has the potential to deliver valuable research insights for studies on student learning, including the important study of physics-problem solving strategies (see Gok, 2010, for a review of research in this area).

We argue that even with modestly sized courses (100-400 students), course-level learning analytics can provide valuable input to instructional designers and instructors, and illustrate this claim by a number of examples from introductory physics courses.

Methodology
The data in this study has been obtained from the log files automatically collected by the LON-CAPA course management system (Kortemeyer et al., 2008). These files record time-stamped and user-identified accesses of course resources such as content pages, homework and exam problems, as well as discussion boards. For problems, in addition the correctness status is recorded for all attempts to solve them. Where applicable, final course grades were also taken into consideration to obtain a broader spectrum of student performance. A number of different

https://doi.org/10.12973/ijpce/17111

introductory physics courses were considered, both algebra- and calculus-based, which are taken as "service courses" by a variety of non-physics majors. Unless mentioned otherwise, these courses took place at Michigan State University. Data were processed using custom scripts and analyzed using correlation and time-series algorithms, as well as classical test and item response theory mechanisms using both custom scripts and the R statistical package (R core team, 2014).

Performance Prediction

As early as 1992, interactive online homework was introduced into undergraduate physics courses (Kashy et al., 1993). Ever since, with about 10 to 20 problems per weekly homework assignment, these courses provide rich performance data. Somewhat unique to physics, these problems tend to be open-ended free-response, where students type in numbers such as "42.6 Nm" – alongside with ranking tasks, these tasks are much more closely coupled with conceptual thinking than simplistic 1-out-of-N multiple-choice problems (Kortemeyer, 2006). Additionally, students typically provide more than one of these complex data points per problem, as multiple tries are allowed to arrive at the correct solution (Kortemeyer, 2015).

In contrast to these rich data, the final result of college courses frequently is just a single grade; not only does the richness of the formative homework data get lost in this measure, but the course grade of a physics course typically has a large number of other components, most notably exams (with usually very high impact on the course grade), labs, clicker points, and recitations. However, homework data comes streaming in from Day One of the course, and in the interest of developing "Early Warning Systems," it is important to ask to what degree Learning Analytics of homework performance can predict the final grade.

We found that up to 87% of the final grade can be predicted by online student behaviors using a combination of different data-mining methods applied to online components of physics courses (Minaei-Bidgoli et al., 2003). As the course grade is largely determined by exam scores, we are essentially investigating how well online behavior predicts exam scores.

The most important predictor is the total number of correctly solved online homework problems (relative importance 100%, independent of how many tries it took), followed by the number of tries required to arrive at the correct solution (relative importance 59%). However, getting a problem correct on the very first attempt is a much weaker predictor (relative importance 28%), with an impact approximately equal to measures of time-spent-on-task (relative importance 25%). Apparently, in physics courses, persistence is a better indicator of success than immediate "genius" (Kortemeyer, 2017). This result is notable, as many physics instructors tend to put more weight on the initial than later attempts, which turns out to be unjustified; it is also surprising, as almost all of the final grade can be explained by data that is plagued by the unproductive student behavior discussed in the following sections. Other online behavior, such as participation in online discussions, was found to be a much weaker predictor of success in our physics courses (relative importance 9%).

Cramming

Traditional physics lecture settings and instructional methods tend to result in ineffective study habits and low conceptual learning (Hake, 1998), as students may not continually reflect on and monitor their learning progress (Mason & Singh, 2009), but instead engage in short "cramming bursts" as exams are approaching; they attempt to catch up and read too much material in too short an amount of time. This behavior cannot be measured directly if the course uses traditional textbooks, as there is no access log. However, self-reporting and analysis of electronic textbooks reveal that continual usage of reading materials is distressingly low in traditional physics courses (Podolefsky & Finkelstein, 2006; Steltzer et al., 2009). This behavior negatively influences course outcomes, since it was shown that interacting with the online materials of physics courses on a daily basis has better learning effectiveness (as measured by exam grades) than only weekly or even less frequent access – continuity matters (Kortemeyer, 2016b).

Learning Analytics can provide insights into the level to which course structure can discourage unproductive behavior (Seaton et al., 2014a). We analyzed the access patterns of online materials in courses at the Massachusetts Institute of Technology and Michigan State University. In courses that were otherwise taught traditionally (weekly end-of-chapter online homework and a small number of midterms), log files show "cramming" before exams, i.e., sharp peaks in access frequency of online materials immediately preceding exams. Despite weekly homework, students do not regularly access the course materials between exams. Access is also highly selective, as students tend to only read materials immediately relevant for their homework, with large portions of course materials only being accessed by a small fraction of the students (Kortemeyer, 2016b).

This unproductive behavior declines when embedding the online homework directly into the materials (as possible in LON-CAPA (Kortemeyer et al. 2008; Seaton et al., 2014b)), paired with frequent exams. We found that while in traditional courses, only half of the students read at least half of the materials, in reformed courses, more than 90% of the students do so – this result is consistent with findings in reformed physics courses elsewhere (e.g., Heiner et al., 2014). The strong access frequency peaks vanish in favor of more distributed smaller peaks. Combined with earlier results that this reformed course structure leads to a more positive attitude and less unproductive problem-solving behavior (Laverty et al., 2012), frequent small quizzes (instead of few large exams) appear to lead to both qualitatively and quantitatively more effective learning environments.

Guessing and Copying

The number of tries used by students to solve online homework problems was found to be a predictor of overall course success (Minaei-Bidgoli et al., 2003). There is no consensus among instructors about how many tries to grant students to arrive at the correct solution; a survey showed that even physics education researchers are granting anywhere from just one to infinite number of tries (Kortemeyer, 2015) and that they also employ different reward schemes including reduced credit for using more tries. These choices matter, as they can influence student behavior when it comes to random guessing or copying of answers (Kortemeyer, 2015). If students are given a large number of attempts, they frequently do not pay much attention to a particular attempt; they tend to enter "random" numbers and work through patterns: changing the sign, changing the result by factors of two or orders of magnitude. We found that more than half of the retries on online homework are submitted within less than one minute of the previous failed attempt when multiple attempts are allowed – hardly enough time for a serious effort on checking answers and derivations (Kortemeyer & Riegler, 2010). Resubmission times are slightly shorter for male than for female students (Kortemeyer, 2009), but in any case below acceptable limits. Granting very few attempts, on the other hand, could lead to students copying answers from others out of apparent desperation.

We analyzed homework transaction logs from introductory physics courses for scientists and engineers (Kortemeyer, 2015), which were taught by different instructors who had different policies regarding allowed tries. We found that the rate at which students succeed on a given attempt decreases with increasing number of allowed tries, and the rate at which students give up and stop working on problems after a given attempt does not depend on the number of allowed tries. We also found that the tries on online homework were largely independent of each other, as students apparently did not learn from their mistakes. This corresponds to the short resubmission times, as well as unsystematic plug-and-chug strategies that do not allow students to go back through their derivations and fix errors (Kortemeyer, 2016a).

Having a large or even unlimited number of allowed tries leads to less desirable problem-solving strategies, as students are not productively taking advantage of these additional chances. Instructors might grant many tries with the intention to allow students to keep working till they master a concept, but we found that running out of tries is a far less likely reason for failure than simply giving up (Kortemeyer, 2015).

A preliminary model, which we developed based on extrapolating trends from the courses with different policies, predicts the optimum number of allowed tries to be five. While students may not be happy with this choice, in reality, it rarely keeps students from eventually solving problems; in a course where five tries were used, the vast majority of students solved almost all homework problems.

Item Response Theory

Our studies show that homework has an essential role in online physics courses. When developing, deploying, and maintaining online courses, their homework components deserve particular attention, and quality control is essential. We have shown that Classical Test Theory can successfully be employed to obtain measures of homework item quality (Kortemeyer et al., 2008), including continual monitoring of these items (Kortemeyer, 2016c)

In summative assessment, e.g., exams, a commonly used tool is Item Response Theory (IRT) (Embredson & Reise, 2013). We studied if this same tool can be used to evaluate the meaningfulness of online homework (Kortemeyer, 2014a) to enhance Learning Analytics. In a possible application of IRT to data mining, online homework items that have more desirable item parameters (i.e., average difficulty and high discrimination) could receive more weight in subsequent data mining efforts (see Baker & Yacef, 2009, for additional avenues to integrating IRT into EDM)

When using IRT to estimate the learner trait of ability, results obtained from online homework show the correct trends compared to exam data, but the absolute agreement is moderate. Our earlier, more classical approaches were more successful to gauge student success (Minaei-Bidgoli et al., 2003). Item discrimination is less sensitive to the change from exam to homework context than item difficulty; while difficulty decreases due to multiple allowed tries and possible copying, homework items still give meaningful feedback to the student and the instructor (Kortemeyer et al., 2005).

Closely related to the above, the effect of considering only the correctness of the first attempt on homework wears out over the course of the semester as the integrity of student study behavior apparently declines: while in the first quarter of the course, it is one of the strongest indicators, over the course of the semester, students start working in less and less desirable ways, and first-attempt success becomes more of an indicator of copying solutions than immediate understanding. As mentioned earlier, eventually solving a problem is a better indicator of student success than getting problems correct on the first attempt.

Attempts to introduce additional student traits in order to lessen the influence of unproductive student behavior such as copying and guessing on item parameters are still an area of active research (Gönülates & Kortemeyer, 2017). Current results, however, indicate that the predictive power of models cannot be greatly improved by introducing additional traits (or, for that matter, additional item parameters); instead, the risk of over-fitting the data increases.

Discussion

The validity of the results from Learning Analytics is overshadowed by the same undesirable student behavior that also hinders learning itself: copying, guessing, and cramming. It was shown that these are reduced by giving more frequent, smaller exams, and by granting a reasonable number of tries. In addition, randomization of online homework can cut down on copying (Kortemeyer et al., 2008). Instructors should also consider types of homework problems that demand more complex solutions than simply numbers, for example graphing tasks (Laverty & Kortemeyer, 2012).

While the mere participation in student discussion is a poor indicator of eventual student success (Minaei-Bidgoli et al., 2003), providing sanctioned online venues within the course context of online courses is a far better option than students migrating to un-sanctioned online "cheat sites" (Kashy et al., 2003). Predictions are, however, possible when considering the actual character of the discussions (Kortemeyer, 2007), and advances in automated online discourse analysis may make it possible to eventually incorporate these characteristics into Learning Analytics.

Finally, it may be a promising approach to add some online problems for which detailed problem derivations need to be submitted. To that end, students can be asked to photograph their written solutions with a smartphone and upload those solutions to the course management system, where they can quickly be hand-graded using a grading queue.

Future applications of Learning Analytics may include recommender systems for instructors, which assist them in compiling course materials and assignments. Of particular interest is mining events where students were unsuccessful in solving a particular homework item, interacted with other materials and problems, and then were successful to solve that item. We expect that given sufficient statistics, these events allow conclusions about the contextual effectiveness of learning resources, rather than considering them in insulation.

Implementing these strategies, unfortunately, is not without challenges: cultural, administrative, and logistical challenges need to be overcome when implementing these methods outside the United States, where many of them originated (e.g., Kortemeyer & Riegler, 2010; Cruz et al., 2011).

Conclusions

Course-level access and formative assessment data provide rich input for Learning Analytics. Results can be used to improve course structure to reduce undesirable behavior, as well as to provide quality control for online materials. Data-Mining methods are overall more promising than traditional Item Response Theory in these online settings, even when attempting to explicitly model unproductive student behavior.

References

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

Baker, R. S. J. D., & 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.

Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX's first MOOC. *Research & Practice in Assessment, 8*.

Bruckman, A. (2006). Analysis of log file data to understand behavior and learning in an online community. In The International handbook of virtual learning environments, 1449-1465.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. G. (2007). Academic analytics: A new tool for a new era. *Educause Review, 42*(4), 40.

Castro, F., Vellido, A., Nebot, À., & Mugica, F. (2007). Applying data mining techniques to e-learning problems. Evolution of teaching and learning paradigms in intelligent environment, 183-221.

Chatti, M. A., Lukarov, V., Thüs, H., Muslim, A., Yousef, A. M. F., Wahid, U., ... & Schroeder, U. (2014). Learning analytics: Challenges and future research directions. eleed, Iss. 10 (urn:nbn:de:0009-5-40350).

Clow, D. (2013). MOOCs and the funnel of participation. In Proceedings of the Third International Conference on Learning Analytics and Knowledge (pp. 185-189). ACM.

Cruz, É., Dias, H., & Kortemeyer, G. (2011). The effect of formative assessment in Brazilian university physics courses. *Revista Brasileira de Ensino de Física, 33*(4), 4315-4315.

Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. In Proceedings of the 1st international conference on learning analytics and knowledge (pp. 9-17). ACM.

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society, 15*(3), 58.

EDUCAUSE Learning Initiative Learning Analytics Resources ... (n.d.). Retrieved July 10, 2016, from https://net.educause.edu/section_params/conf/eli124/LearningAnalyticsResources.pdf

Eichhorn, S., & Matkin, G. W. (2016). Massive open online courses, big data, and education research. *New Directions for Institutional Research, 2015*(167), 27-40.

Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. Psychology Press.

Fournier, H., Kop, R., & Sitlia, H. (2011). The value of learning analytics to networked learning on a personal learning environment. In Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 104-109). ACM.

García Peñalvo, F. J., Conde García, M. Á., Alier Forment, M., & Casany Guerrero, M. J. (2011). Opening learning management systems to personal learning environments. *Journal of Universal Computer Science, 17*(9), 1222-1240.

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends, 59*(1), 64-71.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education, 28*, 68-84.

Gönülateş, E., & Kortemeyer, G. (2017). Modeling unproductive behavior in online homework in terms of latent student traits: An approach based on item response theory. *Journal of Science Education and Technology, 26*(2), 139-150.

Gok, T. (2010). The general assessment of problem solving processes in physics education. *Eurasian Journal of Physics and Chemistry Education, 2*, 110-122.

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society, 15*(3), 42.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics, 66*, 64-74.

Heiner, C. E., Banet, A. I., & Wieman, C. (2014). Preparing students for class: How to get 80% of students reading the textbook before class. *American Journal of Physics, 82*, 989-996.

Kashy, E., Sherrill, B. M., Tsai, Y., Thaler, D., Weinshank, D., Engelmann, M., & Morrissey, D. J. (1993). CAPA—An integrated computer-assisted personalized assignment system. *American Journal of Physics, 61*, 1124-1130.

Kashy, D. A., Albertelli, G., Bauer, W., Kashy, E., & Thoennessen, M. (2003). Influence of nonmoderated and moderated discussion sites on student success. *Journal of Asynchronous Learning Networks, 7*, 31-36.

Khribi, M. K., Jemni, M., & Nasraoui, O. (2008). Automatic recommendations for e-learning personalization based on web usage mining techniques and information retrieval. In Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on (pp. 241-245). IEEE.

Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller, R. C. (2014). Understanding in-video dropouts and interaction peaks inonline lecture videos. Proceedings of the First ACM Conference on Learning @ Scale Conference - L@S '14. ACM.

Kizilcec, R. F., Piech, C., & Schneider, E. (2013, April). Deconstructing disengagement: analyzing learner subpopulations in massive open online courses. In Proceedings of the third international conference on learning analytics and knowledge (pp. 170-179). ACM.

Kortemeyer, G., Hall, M., Parker, J., Minaei-Bidgoli, B., Albertelli II, G., Bauer, W., & Kashy, E. (2005). Effective feedback to the instructor from online homework. *Journal of Asynchronous Learning Networks, 9*(2), 19-28.

Kortemeyer, G. (2006). An analysis of asynchronous online homework discussions in introductory physics courses. *American Journal of Physics, 74*, 526.

Kortemeyer, G. (2007). Correlations between student discussion behavior, attitudes, and learning. *Physical Review Special Topics-Physics Education Research, 3*(1), 010101.

Kortemeyer, G., Kashy, E., Benenson, W., & Bauer, W. (2008). Experiences using the open-source learning content management and assessment system LON-CAPA in introductory physics courses. *American Journal of Physics, 76*, 438.

Kortemeyer, G. (2009). Gender differences in the use of an online homework system in an introductory physics course. *Physical Review Special Topics - Physics Education Research*, *5*(1).

Kortemeyer, G., & Riegler, P. (2010). Large-Scale E-Assessments, Prüfungsvor-und-nachbereitung: Erfahrungen aus den USA und aus Deutschland. *Zeitschrift für E-learning-Lernkultur und Bildungstechnologie*, 8, 22.

Kortemeyer, G. (2014a). Extending item response theory to online homework. *Physical Review Special Topics - Physics Education Research*, *10*(1).

Kortemeyer, G. (2014b). Over two decades of blended and online physics courses at Michigan State University. eleed, Iss. 10. (urn:nbn:de:0009-5-40115)

Kortemeyer, G. (2015). An empirical study of the effect of granting multiple tries for online homework. *American Journal of Physics, 83*(7), 646-653.

Kortemeyer, G. (2016a). The losing battle against plug-and-chug. *The Physics Teacher*, *54*(1), 14-17.

Kortemeyer, G. (2016b). Work Habits of Students in Traditional and Online Sections of an Introductory Physics Course: A Case Study. Journal of Science Education and Technology, 25(5), 697-703.

Kortemeyer, G. (2016c). Scalable continual quality control of formative assessment items in an educational digital library: an empirical study. *International Journal on Digital Libraries, 17(*2), 143-155.

Kortemeyer, G. (2017). The Spectrum of Learning Analytics. eleed, 12(1).

Laverty, J. T., & Kortemeyer, G. (2012). Function plot response: A scalable system for teaching kinematics graphs. *American Journal of Physics, 80*(8), 724.

Laverty, J. T., Bauer, W., Kortemeyer, G., & Westfall, G. (2012). Want to reduce guessing and cheating while making students happier? Give more exams! *The Physics Teacher 50*(9), 540.

Mason, A., & Singh, C. (2009). Reflection and self-monitoring in quantum mechanics. In AIP Conference Proceedings (Vol. 1179, No. 1, pp. 197-200). AIP.

Mattingly, K. D., Rice, M. C., & Berge, Z. L. (2012). Learning analytics as a tool for closing the assessment loop in higher education. *Knowledge Management & E-Learning: An International Journal*, *4*, 236-247.

Mazza, R., Bettoni, M., Faré, M., & Mazzola, L. (2012). Moclog–monitoring online courses with log data, in Proc. 1st Moodle Research Conference, Heraklion, Greece, 132-139.

Minaei-Bidgoli, B., Kashy, D., Kortemeyer, G., & Punch, W. (2003.). Predicting student performance: An application of data mining methods with an educational web-based system. 33rd Annual Frontiers in Education, 2003. FIE 2003.

Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *Journal of Educational Data Mining, 4*, 11-48.

Novak, G. M. (1999). *Just-in-time teaching: Blending active learning with web technology*. Upper Saddle River: Prentice Hall.

Papamitsiou, Z. K., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society, 17*(4), 49-64.

Podolefsky, N., & Finkelstein, N. (2006). The perceived value of college physics textbooks: Students and instructors may not see eye to eye. *The Physics Teacher, 44*, 338-342.

Psaromiligkos, Y., Orfanidou, M., Kytagias, C., & Zafiri, E. (2011). Mining log data for the analysis of learners' behaviour in web-based learning management systems. *Operational Research, 11*, 187-200.

Rayyan, S., Seaton, D. T., Belcher, J., Pritchard, D. E., & Chuang, I. (2014). Participation and Performance in 8.02x Electricity and Magnetism: The First Physics MOOC from MITx. 2013 Physics Education Research Conference Proceedings. comPADRE.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

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

Romero, C., Ventura, S., Delgado, J., & De Bra, P. (2007). Personalized links recommendation based on data mining in adaptive educational hypermedia systems. Creating New Learning Experiences on a Global Scale, 292-306.

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12-27.

Seaton, D. T., Bergner, Y., Kortemeyer, G., Rayyan, S., Chuang, I., & Pritchard, D. E. (2014a). The impact of course structure on eText use in large-lecture introductory-physics courses. 2013 Physics Education Research Conference Proceedings. comPADRE.

Seaton, D. T., Kortemeyer, G., Bergner, Y., Rayyan, S., & Pritchard, D. E. (2014b). Analyzing the impact of course structure on electronic textbook use in blended introductory physics courses. *American Journal of Physics, 82*, 1186-1197.

Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12. ACM.

Stelzer, T., Gladding, G., Mestre, J. P., & Brookes, D. T. (2009). Comparing the efficacy of multimedia modules with traditional textbooks for learning introductory physics content. *American Journal of Physics, 77*, 184-190.

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist, 57*, 1500-1509.

Wolff, A., Zdrahal, Z., Nikolov, A., & Pantucek, M. (2013). Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. In Proceedings of the third international conference on learning analytics and knowledge (pp. 145-149). ACM.

Zaïane, O. R. (2002). Building a recommender agent for e-learning systems. In Proceedings. International Conference on Computers in Education 2002 (pp. 55-59). IEEE.