Relationship between learning engagement metrics and learning outcomes in online engineering course

Tiantian Li
School of Engineering Education
Purdue University
West Lafayette
li1596@purdue.edu

Laura Melissa Cruz Castro
School of Engineering Education
Purdue University
West Lafayette
lcruzcas@purdue.edu

Kerrie Douglas
School of Engineering Education
Purdue University
West Lafayette
douglask@purdue.edu

Christopher G. Brinton
School of Electrical and Computer Engineering
Purdue University
West Lafayette
cgb@purdue.edu

***Abstract*—This research WIP contributes to understanding the relationship between learning engagement in Learning Management System (LMS) and outcomes in an online course. In large engineering courses, it is challenging for instructors to identify who is engaging with course materials at a level necessary to be successful in terms of course outcomes. The purpose of this research WIP study is two-fold: (1) to develop metrics for quantifying learner engagement in online courses, and (2) to explore the relationship between engagement and student success. Our research question is: How does learning engagement relate to course outcomes? We modeled learner engagement on a course level using the following features: number of views per content object, total time spent in the platform, percentage of the course accessed by the learners, percentage of feedback read, and number of attempts per quiz. We used the data collected by the LMS in a large first-year engineering course. We obtained data in Fall 2020, the first semester that many traditional universities were forced mostly or entirely online. After calculating the proposed metrics, we used a linear mixed model to analyze the effect of engagement on learning outcomes. Our linear mixed model shows that all engagement metrics are positively related to the final grade. However, the results also indicate that the relationship between engagement and learning outcomes is not linear; more complex modeling is needed to further explore this relationship.

Keywords—online learning, learning engagement, learning outcome, generalized linear mixed model

I. Introduction

While the role of online learning platforms and learning management systems (LMSs) has been steadily growing, the COVID-19 pandemic has placed online learning front and center in education. LMSs have been an essential part of the daily communication between engineering instructors and students and are an excellent tool for online learning, especially during the pandemic [1]. However, instructors in online courses have noted the difficulty of knowing how students are doing. Compared to in-person teaching, online courses cannot offer instructors chances to gather face-to-face confirmation to ensure the students are on the same page [2]. Given the different educational contexts, instructors need signs or indicators for student engagement in online courses to inform them of students' progress and assist them in providing scaffoldings or diagnostic decisions to facilitate learning.

Online learning platforms, particularly LMSs can help instructors gather information through various learning analytics. For example, data on students' access to feedback can inform the instructors about which feedback has been relevant to their students [3]. Feedback data can also tell the instructor if an intervention has the expected results [4]. The way students watch lecture videos can inform instructors of the challenging materials for their students and guide instructors on providing more support and scaffolding [5]. Other data collected through LMSs, such as number of visits, time spent on the platform, number of discussion posts, etc., has the potential to inform instructors about student engagement with their class and/or the material available on the platform [1].

This paper presents the preliminary efforts of creating metrics for learning engagement for a first-year undergraduate online class and exploring the relationship between learning engagement and learning outcomes. Furthermore, it helps instructors understand student behaviors in online courses and provides a basis for benchmarking online student behaviors.

II. Literature Review

A. Learning Analytics in Learning Management Systems

LMSs are web-based learning platforms that have been widely used in higher education, both in traditional, in-person classrooms and online, virtual instructions as a tool for course administration and pedagogical support [6]. Apart from improving learning and teaching experiences, LMSs also aid in educational research by collecting learning analytics [7], a form of data that can be measured, collected, and analyzed about learners and their learning contexts. Learning analytics can help instructors understand and optimize student learning, especially in an online format [8]. From an instructor's perspective, learning analytics can be used as an additional resource apart from learning outcomes and instruction ratings that are prevalently utilized as course evaluation in most educational settings [9]. Compared to course evaluation, which is usually gathered towards or after the end of a course, learning analytics relays more expedient and automated information to instructors for course improvements [8].

Most LMSs generate basic learning analytics such as activity logs. While the depth of student activity logs differs, student behavior tracking variables usually include the number of visits, time of visits, content visited, and if applicable, the number of posts published on discussion forums, etc. [10]. Some LMSs also collect more detailed clickstream data reflecting learner behaviors such as skip forward, pause, play while watching an embedded video in LMS, and scroll up and down while viewing a PDF or article [11]. Research on learning analytics has utilized this data in various ways, including clustering, relationship mining, prediction and detection of learner behavior, and visualization [8]. Among the various uses mentioned above, one vital application of learning analytics in higher education is identifying latent (i.e., hidden) metrics within the learning process. One such metric is learning engagement, and it has arisen a great interest among the educational analytics community [12].

B. *Engagement*

In the online learning context, engagement is defined as the interest, activity, energy, time, attention, involvement, and other positive characteristics that a learner shows concerning a module, class, content, video, or topic [13]–[15]. Engagement is a predictor of student retention and success, and it is understood to be an indicator of the quality of student experience for higher education [6], [14], [16]. Although it is possible to confuse engagement with an observable construct, engagement is latent [17]. It is unobservable and therefore only approximately measured. Different theories are used to find the composition of engagement. One theory, used in the National Student Engagement Survey (NSES), describes the composition of engagement in four factors: skills engagement (i.e., staying up on readings, putting forth effort), emotional engagement (i.e., making the course enjoyable, applying it to my life, etc.), participation/interaction engagement (i.e., having fun, participating actively in small group discussions), and performance engagement (e.g., doing well on tests, getting a good grade) [18].

There are at least three sources of information that are prominent in literature: (1) using instruments to collect self-reported information (e.g., NSEE [15], thematic analysis of surveys [19], (2) using external technological devices to capture neural information (e.g., Emotient, SightCorp [20]), and (3) clickstream activity-based data (e.g., LMS information [12]). Among the three sources of information, LMS data may become one of the most used because of its wide availability [21]. Engagement with learning nowadays likely means engagement with technology [22]. Therefore, most institutions have started to rely on LMSs to support student-instructional team communication. These online platforms have become an additional source of information that helps understand what students do outside of the classroom [16], [23]. Engagement within the LMS context is often measured by the number of followers, likes, discussions, comments, and shares, and the duration of interaction in the online domain [24]. Some methodologies have used data analysis to generate an engagement score [17]. In contrast, others have found some metrics related to engagement factors and used them to create predictive models of learning outcomes [11], [12], [25].

In the current study, we want to explore different metrics that can be easily obtained and that have the potential of measuring the interaction/participation engagement of students within the LMS context. Using these metrics, we want to answer the research questions: *How does learning engagement relate to course outcomes in an online setting?*

III. METHODS

A. *Data Collection and Context of the Study*

We used the data collected by the LMS for a first-year engineering course at a large public university. The institution uses Brightspace as its LMS. The first-year engineering course is the first required engineering course for most first-year engineering students. The main learning objectives of this course include introducing students to mathematical modeling, the design process, and making evidence-based engineering decisions while collaborating in diverse teams. The data contains information on enrollment and withdrawal, assignment submissions, quiz attempts, course accesses, and grade results. The data set corresponds to over 1700 students enrolled in this course in Fall 2020, across 17 sections. The university's engineering freshman cohort in Fall 2020 consists of 26.33% female students and 74.67% male students. Among these students, 55.67% identified as White, 13.29% identified as Asian, and 1.57% identified as Black or African American [26]. During this semester, instructions were moved into an online format to comply with the COVID-19 restriction.

B. *Data Availability & Metrics for Engagement*

After obtaining the data, we started designing and computing metrics that could be used to represent learning engagement. Drawing from previous studies that quantified learning engagement while accommodating the available information our data offers, we went through an iterative process to generate and define learning engagement metrics on an overarching level. The following metrics were the results of this process:

1) Views per content object (VC): the total number of views where a content object is viewed divided by the total number of content accessed by students within the section.

2) Total time spent in real visits (TT): the total time spent in all the real visits (a visit where content is viewed and the user continues in Brightspace [27]) within the section, measured in hours.

3) Percentage of course accessed (CA): the number of content visited by the student divided by the total number of content accessed by students within the section.

4) Percentage of feedback read (FR): the number of feedback read by students divided by the total number of feedback from individual assignments within the section.

5) Attempts per quiz (QA): the total number of quiz attempts made by student divided by the number of quizzes within the section.

Finally, we obtained the final grade allocated in a percentual scale for each student. In the first-year course, the final grade was a combination of design project (40%), quizzes and two mid-term exams (37%), teamwork (10%); the remaining 14% of the grade is distributed between preparation before classes, in-class participation, etc. The final grade percentage was calculated in an unadjusted manner.

C. *Generalized linear mixed model*

We hypothesize that the five metrics defined in Sec. III-B are different dimensions of engagement, though the exact formula for translating to a single engagement score is unclear. In addition, we consider the interpretability of the results an important part of this study since it could help understand the role of each dimension on the learning outcomes. Therefore, we use a Generalized Linear Mixed Model (LMM) to analyze the effect of engagement, as defined in this study, on students' final grades [28]. A LMM is a model calculated with a Gaussian link, that in addition to the fixed effects that are usually added in a linear model (e.g., linear regression model), contains a random variance parameter that accounts for the variance explained by the sampling hierarchical structure, allowing a better generalization. The effect of the instructional teams, as well as other features of the group in which they were assigned,.1)

$$\hat{y} = \beta X + \gamma Z \quad (Eq. 1)$$

IV. RESULTS

The descriptive statistics of high-level metrics we generated to evaluate student engagement in the first-year engineering course, together with the final grade percentage, are listed in Table 1. Final grades in this first-year engineering course have a mean of 92.65 out of 100 and an SD of 6.65.

TABLE I. DESCRIPTIVE STATISTICS OF ENGAGEMENT METRICS & FINAL GRADES

Metrics	*M*	*SD*	Min	Max
Views per content object (VC)	0.84	0.24	0.12	1.93
Total time of real visits (hr) (TT)	29.79	20.24	2.45	320.84
Course accessed (%) (CA)	45.52	10.24	11.07	92.37
Feedback read (%) (FR)	6.22	7.80	0	80.39
Attempt per quiz (QA)	1.37	0.22	0.23	2.56
Final grade (%)	92.63	6.65	26.62	100.00

Based on the correlation matrix in Fig. 1, we can see that among the learning engagement metrics, the percentage of content access and the number of views per content object are strongly correlated with each other, having a correlation coefficient of 0.82. Surprisingly, neither the percentage of feedback read nor the number of attempts per quiz is highly correlated with total time spent in real visits. Due to the high correlation between CA and VC, VC was not included in the scatter plots between engagement metrics and learning outcomes in Fig. 2. However, we did not discard the metrics in future descriptive analysis since they represent distinct dimensions of learning engagement.

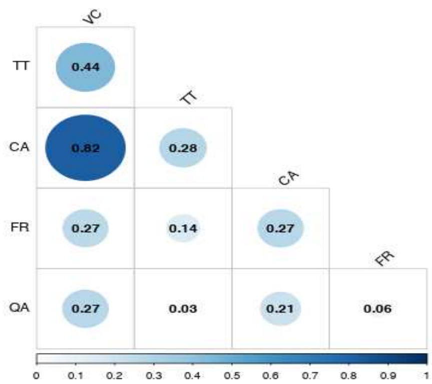

Fig. 1 Correlation matrix of high-level learning engagement metrics. Variable description: VC = views per content object, TT = total time spent in real visits in hours, CA = percentage of content accessed, FR = percentage of feedback read, and QA = attempt per quiz.

We descriptively evaluated the relationship between final grades and the proposed metrics and built scatter plots between the metrics and the learning outcomes to gather more information about the relationships (Fig.2). We used a logarithm scale on the x-axis to better visualize the data points and to account for influential points existing in the data. The scatter plots were helpful in concluding that the relationships between the metrics that we have defined and the learning outcomes could be linear in some regions, but a linear relationship might not be sufficient to express the relationship in the entire area. In addition, the metrics selected seem to have a positive relationship with learning outcomes.

Our linear mixed model results in Table 2 showed that

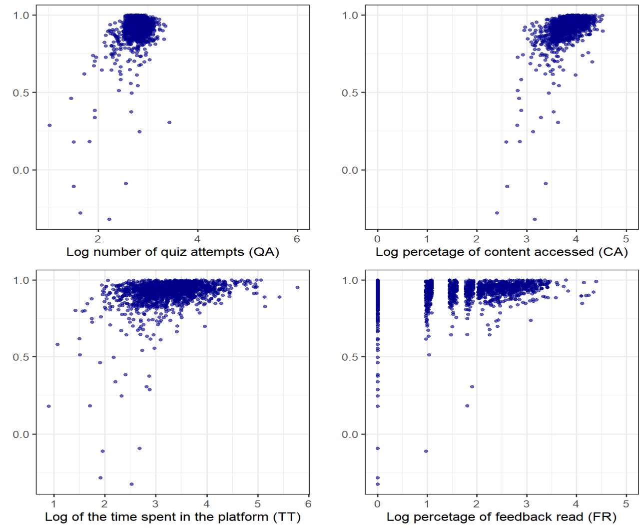

Fig. 2 Scatter plots of engagement metrics and learning outcomes.

42.05% of the variance in final grades could be explained by learning engagement metrics when fitting the variables into a simple linear regression. All metrics are significant predictors of final grades. All of the engagement metrics are also positively correlated with the final grades. However, some significant interaction terms are negatively correlated with final grades. These terms are, the interaction of the number of attempt per quiz and percentage of course accessed, the interaction of the total time of real visits and percentage of feedback read, and interaction of percentage of course accessed and percentage of feedback read.

TABLE II. SUMMARY OUTPUT OF LINEAR MIXED MODEL

Variables	Estimate	Standard error	T value
Intercept	25.99	2.50	10.43*
Number of quiz attempts	3.05	0.20	19.15*
Percentage of content accessed	1.43	0.06	23.40*
Percentage of feedback read	0.49	0.07	6.62*
Total time of real visits	0.16	0.03	5.22*
Number of quiz attempts <> Percentage of content accessed	-0.07	0.0004	-17.45*
Total time of real visits <> Percentage of feedback read	-0.0008	0.00006	-4.74*
Percentage of content accessed <> Percentage of feedback read	-0.0003	0.0012	-5.90*

Note: * indicates significance at a 0.05 or lower level. <> indicates an interaction between two variables

V. DISCUSSION

Our mixed linear model explains 42% (R-square) of the variance in final grades, showing a partially linear relationship between learning engagement as defined by the metrics and final grades. Our linear model and learning engagement metrics confirm the findings on some previous literature revealing an overall positive relationship between LMS usage and course grades. Similar previous studies revealed that higher LMS usage, such as the number of content objects read, could lead to better learning outcomes [18], [16]. In particular, a study using eight participation measures, including time spent and the number of contents viewed on different parts of the LMS, reported 31% of the variance in student success [29]. Since the model explained 42% of the variance in final grades, the engagement metrics proposed can be a potential tool for instructors to monitor student learning using LMS data in online courses.

Apart from the overall positive trend between engagement and outcomes, this study also finds that the relationship between engagement and learning outcomes is more complex than what we hypothesized before. In other words, our model shows that a linear monotonic trend between engagement and learning outcomes would limit the usability of key metrics such as percentage of content access. For example, it was found that a lower grade automatically guarantees a low percentage of content accessed while a low percentage of access does not ensure a low final grade. One possible explanation for this phenomenon is that the percentage of content access data might contain information on the importance of accessing more specific content. Therefore, a more strategic student would access certain materials and succeed in acquiring the information needed. Another indicator of a more complex relationship between engagement and outcomes is the discrepancy between individual- and group-based portion within engagement and final grades in our data. In the engagement metrics, we excluded all the data from group-based activities in the course. However, the same measure cannot be repeated with learning outcomes, as the final grade includes 10% on teamwork and several group design projects. As a result, students' grades were a combination of individual and group efforts. Metrics for individual learning outcomes need to be generated in future studies to filter out the noise introduced by teamwork. Combined with the trends between these learning attributes shown in Fig. 2, a more complex model should be explored in future research to develop a better model to predict learning outcomes based on the learning analytics that represent learning engagement.

The findings show that total time spent on the platform is the least significant predictor among all engagement metrics. This result differs from the finding of [29], in which the time spent (viewing discussion posts) was found as one of the main predictors. This discrepancy can be due to the different ways of recording time spent on the platform by different LMSs and the noise introduced when web browser windows are not being closed by users. In the case of this study, Brightspace defines views into two kinds: real and fake. They are distinguished by whether subsequent activities were generated by the user or not. Only the time spent on real views is recorded. Information on instances where a student viewed content and proceeded to use other tools such as design software before being automatically logged out of Brightspace due to inactivity was lost. Therefore, depending on how students utilize the LMS, the total time spent might differ in its ability to predict final grade.

In conclusion, our findings support the literature on learning analytics, specifically, the importance of learning analytics to understand and optimize student learning [10]. Unlike the traditional classroom settings, online courses do not offer instructors the chance to observe student behaviors first-hand. In some online courses, instructors even lose the opportunity to gauge engagement through attendance. Since the assessment of engagement using learner analytics involves a combination of factors and various kinds of student behavioral data, the evaluation of online courses should also examine all the metrics related holistically to help instructors provide scaffolding to learners exhibit different engagement behaviors.

VI. Limitations

Four limitations are highlighted for this study: (1) This study defined learning engagement in terms of metrics collected only from LMS, making the learning analytics platform-dependent. In the case of our study, we had access to learner’s access logs on an aggregated level (e.g., data on login sessions and clickstreams during each login session was unavailable), making the metrics generation process more challenging. Therefore, we were unable to conduct validations and optimize our metric selection. (2) Due to the limitations of the platform, the time dimension of the data was not present in our analysis. (3) This study only used linear models instead of more complex models and techniques to understand the relationship between learning engagement metrics and learning outcome. (4) The current sample was limited in its diversity. Our sample showed a small number of female students and non-white students. In addition, gender was collected in a binary manner, further limiting our demographic knowledge.

VII. Conclusion

In conclusion, it was possible for us to generate five course-level metrics to latently represent learning engagement in an online first-year engineering course. The metrics are: the number of views per content object, total time spent in real visits, percentage of content accessed, percentage of feedback read, and the number of attempts per quiz. All metrics are positively correlated with students' grades. The linear mixed model showed a positive relationship between engagement and learning outcomes. With the results of our linear mixed model, we offer a potential way for online course instructors to gauge students' engagement level by learning analytics metrics obtained from the LMS in the absence of face-to-face confirmation of student comprehension and learning progress. Our results indicate that analysis of engagement and student success for online courses should be done on a holistic level, examining all factors related to student participation. Using the proposed metrics, instructors can provide scaffoldings and interventions to encourage and facilitate engagement, thus promoting student success in online courses. Our future work includes exploring non-linear modeling methodologies to increase the accuracy (i.e., R-square) of our model, performing clustering analysis to see whether patterns of student engagement behaviors emerge, examining the relationships between engagement patterns and learning outcomes, and generating dashboards that contain visualizations to help instructors monitor student engagement in real time and offer scaffolding.

Acknowledgment

This work is possible from funding from Koch Family Foundation. We want to thank Leyla Ciner for her review of the document, and Beth Tenorio for help provided on data extraction and suggestions made to improve the data analysis.

References

[1] P. Paudel, "Online education: Benefits, challenges and strategies during and after COVID-19 in Higher Education," *Int. J. Stud. Educ.*, vol. 3, no. 2, pp. 70–85, 2020.

[2] M. Rannastu-Avalos and L. A. Siiman, *Challenges for distance learning and online collaboration in the time of covid-19: Interviews with science teachers*, vol. 12324 LNCS. Springer International Publishing, 2020.

[3] J. Li, S. C. Wong, X. Yang, and A. Bell, "Using feedback to promote student participation in online learning programs: evidence from a quasi-experimental study," *Educ. Technol. Res. Dev.*, vol. 68, no. 1, pp. 485–510, 2020.

[4] M. D. Pistilli, "Learner Analytics and Student Success Interventions," *New Dir. High. Educ.*, vol. 2017, no. 179, pp. 43–52, 2017.

[5] L. Angrave, Z. Zhang, G. Henricks, and C. Mahipal, "Who benefits? Positive learner outcomes from behavioral analytics of online lecture video viewing using classtranscribe," *Annu. Conf. Innov. Technol. Comput. Sci. Educ. ITiCSE*, pp. 1193–1199, 2020.

[6] H. Coates, R. James, and G. Baldwin, "A critical examination of the effects of learning management systems on university teaching and learning," *Tert. Educ. Manag.*, vol. 11, pp. 19–36, 2005.

[7] L. P. Dringus and J. A. Scigliano, "From early to current developments in online learning at Nova Southeastern University: reflections on historical milestones," *Internet High. Educ.*, vol. 3, no. 1–2, pp. 23–40, 2000.

[8] J. T. Avella, M. Kebritchi, S. G. Nunn, and T. Kanai, "Learning analytics methods, benefits, and challenges in higher education: A systematic literature review.," *Online Learn.*, vol. 20, no. 2, pp. 13–29, 2016.

[9] M. S. Medina, W. T. Smith, S. Kolluru, E. A. Sheaffer, and M. DiVall, "A Review of Strategies for Designing, Administering, and Using Student Ratings of Instruction," *Am. J. Pharm. Educ.*, vol. 83, no. 5, p. 7177, Jun. 2019.

[10] L. P. Macfadyen and S. Dawson, "Mining LMS data to develop an 'early warning system' for educators: A proof of concept," *Comput. Educ.*, vol. 54, no. 2, pp. 588–599, 2010.

[11] W. Chen, C. G. Brinton, D. Cao, A. Mason-Singh, C. Lu, and M. Chiang, "Early Detection Prediction of Learning Outcomes in Online Short-Courses via Learning Behaviors," *IEEE Trans. Learn. Technol.*, vol. 12, no. 1, pp. 44–58, 2019.

[12] A. S. Lan, C. G. Brinton, T. Y. Yang, and M. Chiang, "Behavior-based latent variable model for learner engagement," *Proc. 10th Int. Conf. Educ. Data Mining, EDM 2017*, pp. 64–71, 2017.

[13] C. R. Henrie, R. Bodily, R. Larsen, and C. R. Graham, "Exploring the potential of LMS log data as a proxy measure of student engagement," *J. Comput. High. Educ.*, vol. 30, no. 2, pp. 344–362, 2018.

[14] P. Redmond, L. A. Abawi, A. Brown, R. Henderson, and A. Heffernan, "An online engagement framework for higher education," *Online Learn. J.*, vol. 22, no. 1, pp. 183–204, 2018.

[15] C. C. Robinson and H. Hullinger, "New Benchmarks in Higher Education: Student Engagement in Online Learning," *J. Educ. Bus.*, vol. 84, no. 2, pp. 101–109, 2008.

[16] S. Rafaeli *et al.*, "OnLine, Web Based Learning Environment for an Information Systems course: Access logs, Linearity and Performance," no. May 2014, 1997.

[17] V. Singh *et al.*, "A content engagement score for online learning platforms," *Proc. 5th Annu. ACM Conf. Learn. Scale, L S 2018*, pp. 1–4, 2018.

[18] C. Brozina, D. B. Knight, T. Kinoshita, and A. Johri, "Engaged to Succeed: Understanding First-Year Engineering Students' Course Engagement and Performance through Analytics," *IEEE Access*, vol. 7, pp. 163686–163699, 2019.

[19] R. Cobos and J. C. Ruiz-Garcia, "Improving learner engagement in MOOCs using a learning intervention system: A research study in engineering education," *Comput. Appl. Eng. Educ.*, no. January, pp. 1–17, 2020.

[20] A. Kamath, A. Biswas, and V. Balasubramanian, "A crowdsourced approach to student engagement recognition in e-learning environments," in *2016 IEEE Winter Conference on Applications of Computer Vision, WACV 2016*, 2016.

[21] E. W. Black, K. Dawson, and J. Priem, "Data for free: Using LMS activity logs to measure community in online courses," *Internet High. Educ.*, vol. 11, no. 2, pp. 65–70, 2008.

[22] G. Stahl, "Foreword," in *Engaged learning with emerging technologies*, D. Hung and M. S. Khine, Eds. Dordrecht, NL: Springer, 2006.

[23] H. A. . Diefes-Dux and L. Cruz, "Student reflection to improve access to standards-based grading feedback," in *Frontiers in Education 2018 - Fostering Innovation Through Diversity*, 2018, p. 83.

[24] W. Wang, L. Guo, L. He, and Y. J. Wu, "Effects of social-interactive engagement on the dropout ratio in online learning: insights from MOOC," *Behav. Inf. Technol.*, vol. 38, no. 6, pp. 621–636, 2019.

[25] C. G. Brinton, S. Buccapatnam, M. Chiang, and H. V. Poor, "Mining MOOC Clickstreams: Video-Watching Behavior vs. In-Video Quiz Performance," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3677–3692, 2016.

[26] Purdue University, "Student Enrollment," *Purdue Data Digest*. .

[27] Brightspace, "Brightspace Data Sets for Content," *Brightpace Help*. .

[28] C. E. Mcculloch and S. R. Searle, "GLMMs," in *Generalized, linear, and Mixed models*, New York, NY: John Wiley & Sons, 2001, pp. 220–247.

[29] L. V Morris, C. Finnegan, and S.-S. Wu, "Tracking student behavior, persistence, and achievement in online courses," *Internet High. Educ.*, vol. 8, no. 3, pp. 221–231, 2005.