The validity and utility of activity logs as a measure of student engagement

Benjamin Motz
Department of Psychological and Brain Sciences, Indiana University, Indiana, United States
bmotz@indiana.edu

Joshua Quick
eLearning Design and Services, Indiana University, Indiana, United States
jdquick@iu.edu

Noah Schroeder
eLearning Design and Services
Indiana University, Indiana, United States
schroedn@iu.edu

Jordon Zook
eLearning Design and Services
Indiana University, Indiana, United States
jpzook@iu.edu

Matthew Gunkel
eLearning Design and Services
Indiana University, Indiana, United States
mgunkel@iu.edu

ABSTRACT

Learning management system (LMS) web logs provide granular, near-real-time records of student behavior as learners interact with online course materials in digital learning environments. However, it remains unclear whether LMS activity indeed reflects behavioral properties of student engagement, and it also remains unclear how to deal with variability in LMS usage across a diversity of courses. In this study, we evaluate whether instructors' subjective ratings of their students' engagement are related to features of LMS activity for 9,021 students enrolled in 473 for-credit courses. We find that estimators derived from LMS web logs are closely related to instructor ratings of engagement, however, we also observe that there is not a single generic relationship between activity and engagement, and what constitutes the behavioral components of "engagement" will be contingent on course structure. However, for many of these courses, modeled engagement scores are comparable to instructors' ratings in their sensitivity for predicting academic performance. As long as they are tuned to the differences between courses, activity indices from LMS web logs can provide a valid and useful proxy measure of student engagement.

CCS CONCEPTS

• Applied computing ~ Learning management systems
• Information systems ~Web log analysis

KEYWORDS

LMS, web logs, student engagement, trace data

ACM Reference Format:

B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. 2019. The validity and utility of activity logs as a measure of student engagement. In *The 9th International Learning Analytics & Knowledge Conference (LAK19), March, 2019, Tempe, AZ, USA*, 10 pages. https://doi.org/10.1145/3303772.3303789

LAK19, March, 2019, Tempe, AZ, USA

ACM 978-1-4503-6256-6/19/03…$15.00
https://doi.org/10.1145/3303772.3303789

1 INTRODUCTION

When a student downloads an assigned reading, reviews the instructions for an upcoming assignment, prepares a submission, and ultimately turns in school work, all these activities are timestamped and logged by default in contemporary learning management systems (LMSs). These activity logs divulge student behaviors that, until recently, might have been completely unobservable to teachers and advisors, traditionally carried out in the privacy of dorm rooms, libraries, and local coffee shops. Previously, to assess a student's level of engagement, a teacher would need to reflect on the student's submitted work and in-class participation – however, when a student's clicks on learning resources, readings, and course assignments are being logged moment-by-moment, could these activity logs provide a valid proxy measure of the student's engagement? The goal of the current study is to address this question by investigating the relationship between LMS activity logs and instructors' own assessments of student engagement at an institution-wide scale.

Recent research has revealed that, in many isolated contexts, the quality and quantity of student activity in the LMS are positively related to student performance (see Section 1.1). However, there is a clear difference between identifying the activity of successful students and identifying the behaviors that are characteristic of engaged learning. For example, if the activity patterns of successful students reflect dispositional properties of the individual (e.g., successful students happen to use the LMS more), the activities observed from LMS logs are no more actionable than students' past grades. Additionally, one could easily imagine reasons that LMS activity might be unrelated – or inversely related – to student engagement, even despite the positive association between activity logs and student performance. Just as increased social media usage can correlate with withdrawal from ones' social relationships [1], it is possible that increased LMS usage might belie deficiencies in classroom engagement. For example, perhaps disengaged students who sleep through class will compensate by increasing their utilization of online resources; perhaps students' levels of interaction in an online learning environment reflect more general inclination toward technology tools, and not toward engagement within the educational context; or perhaps students who tend to leave browser windows open, click aimlessly on course resources, and interact haphazardly with an LMS *also* just happen to be successful students. For these reasons, we see a unique need to assess the relationship between LMS activity and student engagement across a very wide sample of courses and students, particularly if actionable intelligence is to be derived from these data sources [2].

1.1 Activity Logs

Rapidly increasing utilization of LMSs [3] has sparked increased interest in the detailed web logs unobtrusively recorded within these systems describing student activity. At this time there is no clear consensus on precisely what kinds of data are encompassed under the umbrella term "activity logs," as these may sometimes include very granular markers of events and milestones within a single learning object (e.g., IMS Global's *Caliper*), or these activity logs may be comprised of more coarse measures of browser navigation events as students transition across different screens within the platform (as in the current study). There is also no consensus on how, precisely, to quantify or analyze these web logs, but nevertheless, a wide variety of indicators derived from LMS activity logs have been reported to be reliably correlated with course performance across a range of isolated samples [4-14].

This positive relationship between LMS activity and course performance offers great promise as a means of identifying students who are at risk of failure or dropout from a course [15]. Considering that student activity within a course is directly causally related to student success within that course, LMS activity indicators might be uniquely actionable as a way to identify such risk *proactively*, before the student receives a negative grade consequence [9, 16-18]. Whereas the initial wave of predictive learning analytics focused on predicting future performance from past performance (which is outside a student's immediate control), LMS activity logs at least have the potential to reveal problems in student engagement that are directly remediable.

Optimism about the utility of activity log data is buttressed by researchers' confident claims that learners' navigation patterns reflect self-regulated studying trends and strategic resource usage [9, 10, 19-21]. Indeed, proactive interventions specifically aimed at increasing interaction within the LMS seem to have positive benefits for student performance [22, 23]. Taken together, one might be led to believe that activity logs measure a general theoretical construct akin to student engagement and that individual differences in these activity logs can augment predictive models of student success.

However, Conjin et al. [9] recently raised two concerns necessitating caution about the validity and utility of broad activity indices as a measure of engagement. First, the authors pointed out that, while LMS activity is correlated with course performance, it remains unclear what these logs are measuring about the student mindset. While many learning analytics researchers assume that LMS activity is indicative of engaged learning, its construct validity has not been directly examined. Second, the authors noted that LMS activity is strongly determined by the structure of the course and the instructor's use of online tools within this structure (see also [12, 24]). In this regard, modeled indices of activity logs are not easily "portable" across courses, potentially impeding the use of such indicators at scale (see also [25]).

The difficulty in enabling model transfer across contexts reflects a central tension between generalized and context-specific models in learning analytics. Gasevic et al's [24] findings on the limitations of more generalized models for predicting student performances raises pertinent questions on the relevance and applicability of using LMS activity logs, or any student behavior, for predictive analytics without also considering the instructional and pedagogical conditions of the courses being modeled (see also [26]). Keeping these tensions in mind, we sought to identify common groupings of LMS tool usage within activity logs (see Section 2.3). We view this approach as a movement toward a middle ground between generalized and context-specific predictive models by acknowledging the existence of differences in LMS activity and tool use across course contexts but also understanding that different students may exhibit some similarities in their general activity patterns within LMS web logs that can lead towards more scalable applications.

1.2 Student Engagement

Student engagement represents a contentious theoretical concept within educational research. To the extent that there is consensus in the literature, researchers and theorists agree that engagement is a multifaceted, complex construct [27, 28]. In educational contexts, engagement may include affective states (e.g., emotions and attitudes about the learning task), individual dispositional states (e.g., goals and motivation for learning), and specific behaviors in situ, at the moment of an educational activity (e.g., action and effort during learning). The former two factors (affective and stable individual states) are certainly relevant to a comprehensive formulation of student engagement, but they are also operationally elusive – difficult to define and measure in practice – and they are also dynamically interdependent with learner actions. Similarly, the aspects of student engagement that are specific to social situations and academic disciplines are difficult to observe and identify at larger scales without losing the contextual significance of the student's situated learning [29].

For these reasons, we join other researchers in focusing primarily on specific *behavioral* aspects of engagement [30, 31, 32]. Importantly, we do not claim that this represents a complete picture of a student's engagement in educational contexts. We do not propose to validate LMS records as measuring a student's eagerness to raise her hand in an accounting class, or a student's intrinsic motivation to master a logical proof, both of which are interesting, relevant, and important aspects of a broad definition of engagement. We do, however, aim to examine the validity of LMS records as measuring specific behavioral dimensions of engagement, such as attending, participating in, and completing coursework for their enrolled classes.

For judging these behavioral dimensions of student engagement, teachers have privileged insights because they are expressly responsible for setting the norms of learning behaviors within the unique contexts of their courses. For example, the teacher determines the extent to which class attendance represents a meaningful element of student behavior in their course, and thus the teacher's personal appraisal of student engagement regarding attendance will be more precise than any generic formulation. Research has demonstrated that teachers' estimates of their students' levels of engagement along such behavioral dimensions are accurate [33]. Further, teachers' ratings of student engagement within any given course context reflect both situated and social components of engagement embedded within their pedagogy and disciplinary practices. These ratings are therefore indicative of aspects of student engagement that are not readily apparent directly from the behavioral components recoverable from activity logs. Consequently, we frame the current study as identifying valid techniques for measuring specific behavioral aspects of engagement, as judged by teachers, from LMS activity logs.

1.3 The Current Study

The current study examines the construct validity of activity logs as a measure of student engagement. Specifically, we investigate the relationship between features of student activity derived from LMS web logs, and instructors' ratings of student engagement. As many analysts presume, it is possible that activity logs provide a valid measurement of engaged student behavior as assessed by the instructor. However, it is also possible that LMS activity logs can reflect systematic characteristics of *dis*engagement; for example, perhaps students who are skipping class compensate by accessing more online resources, submitting more assignments online, or otherwise making heavier use of the LMS; and if this is the case, we would not expect any strong positive overall correlation between instructors' assessments of student engagement and LMS activity features.

To examine these possibilities, we capitalize on a local institution-wide dataset containing teachers' ratings of their students' engagement across hundreds of courses and thousands of students, spanning nine Indiana University campuses. As part of its student success initiatives, IU encourages faculty to manually submit a *Student Engagement Roster* for each of their courses, raising flags on behavioral dimensions of engagement for their enrolled students. We consider these ratings as "ground truth," providing authentic natural classifications of students according to their levels of engagement. Our goal is to examine the relationship between these ratings, and features of student activity derived from LMS web logs.

Even so, it would be untenable to assume that there exists one single relationship between student activity and instructor ratings of engagement. Structural, social, and other contextual differences between courses will affect how students behave in the course sites, and these will also mediate how student behavior correlates with engagement. Rather than subjectively coding these differences between courses, we sought to classify these differences directly from student activity itself. Specifically, we used a straightforward clustering technique to classify courses according to the web logs of typical students in each of these courses. From this, we analyzed the relationship between activity and engagement separately for each cluster using logistic regression, with the aim of examining the correlation between students' relative levels of activity within the LMS, and instructors' subjective ratings of those students' levels of engagement in the course.

Considering that the popular utility of activity logs is to ultimately identify students who are at risk of poor academic performance, we also sought to investigate the sensitivity of modeled estimates of student engagement for predicting negative grade outcomes, and how this sensitivity compares with instructors' ratings. Rather than fitting models to predict grade outcomes directly, the goal of this analysis is to assess whether the variance in LMS activity logs specifically associated with *student engagement* is also a valid predictor of student success.

All analytical scripts for the current study are publicly available on OSF at https://osf.io/bt3xn/.

2 METHOD

2.1 Sample

All Indiana University class sections in which the instructor submitted engagement ratings (of any kind, for any student; see Section 2.4, below) were identified during the Fall 2017, Spring 2018, and Summer 2018 semesters. There were 997 such class sections, managed within 829 distinct LMS course sites (some sections were cross-listed within a single parent site, others had no course sites). We then excluded course sites containing students with multiple graded enrollments or enrolled on a non-graded basis ($n = 28$), containing enrolled graduate students (remaining $n = 71$), and containing fewer than 10 enrolled students (remaining $n = 257$), yielding 473 total course sites (with 11,926 total enrollments from 9,021 unique students) for the current analysis. Retrospective analysis of engagement records and LMS activity logs from these courses was approved by the Indiana University Institutional Review Board.

2.2 Feature Extraction

Features of student activity (such as the amount of time a student spent on assignment pages in a course site) were extracted directly from the raw web logs of browser-based navigation in Canvas (Instructure; Salt Lake City, Utah), our institution's LMS, for each student in each course site. For the current analysis, we chose 19 features we believed would be relevant to a student's engagement with online assignments specifically (and not, for example, online quizzes or online discussions), because 'Assignments' is the most widely-used LMS tool at our institution and the primary form of student interaction within Canvas. See Table 1 for a list of features; our code used for extracting these features from Canvas's web logs is publicly available at https://osf.io/ghsbf/.

Features were extracted for the time period beginning at the official start date of each class, and up to when the instructor submitted the earliest rating of a student's *Overall Engagement* or

	Cluster					
	1	2	3	4	5	6
***Descriptive Statistics** (not used in clustering)*						
Number of courses in cluster	16	6	12	194	198	47
Avg course enrollment	24.3	37.7	19.8	27.5	24.2	20.1
Unique students in cluster	384	224	236	4,698	4,124	929
Total negative engagement flags raised	73	7	13	597	488	146
Percent of courses that are hybrid or online	31.3%	0	25.0%	44.8%	8.6%	53.2%
Features of Student Activity						
Time on asgmt pages (m)	35.5	49.2	896.7	57.5	13.3	371.9
Avg time between first access & asgmt deadline (h)	391.3	1,556.9	35.6	54.3	32.8	65.1
Avg session duration with asgmt views (h)	2.5	3.0	4.2	2.5	1.7	3.8
Avg page views / session with asgmt views (c)	14.5	16.0	11.8	14.5	11.7	16.6
Visits to 'Files' after an asgmt view (c)	1.1	0.3	2.1	1.3	0.7	1.5
Visits to other 'Assignments' after an asgmt view (c)	2.8	3.1	19.8	3.3	1.5	8.6
Visits to 'Modules' after an asgmt view (c)	0.2	0.0	1.2	0.4	0.1	2.6
Visits to static 'Pages' after an asgmt view (c)	0.1	0.0	0.2	0.1	0.0	0.4
Total asgmt views with no subsequent visit (c)	5.6	6.5	29.9	5.9	3.0	16.5
Visits to other Canvas tools after an asgmt view (c)	1.0	1.9	5.0	2.8	1.0	4.3
Number of asgmt submissions 6am-6pm (c)	1.2	1.3	5.8	2.5	1.2	5.0
Number of asgmt submissions 6pm-midnight (c)	0.1	0.1	0.8	0.4	0.1	0.8
Number of asgmt submissions midnight-6am (c)	0.8	0.8	4.5	1.7	0.6	3.8
Total number of submissions (c)	2.1	2.3	12.3	4.8	2.2	10.3
Total visits to asgmt pages before deadline (c)	10.8	5.4	53.2	14.5	5.3	35.6
Total visits to asgmt pages after deadline (c)	1.4	0.0	15.7	2.9	1.3	9.2
Number of unique sessions with site visits (c)	11.2	11.9	25.4	11.6	10.3	19.4
Visits to Canvas's 'Calendar' of assignments (c)	3.4	5.0	2.6	2.6	2.6	2.9
Longest period of inactivity within the site (h)	145.6	156.1	154.3	92.2	191.6	118.8

Table 1: Course clustering based on student activity. 'asgmt' is short for assignment; (h) indicates hours, (m) indicates minutes, (c) indicates count. Values shown are descriptive statistics for courses assigned to each cluster, and cluster centroids for features of student activity. Centroids describe the tendency of the associated courses' median values (across students), for each feature.

their engagement with *Assignments* (see 2.4, below). In cases when an instructor did not submit any rating along these two dimensions (such as instructors who only submitted Attendance ratings), we used the average rating date for these dimensions in other courses within the respective academic term.

Features were measured for each student in each course site. Missing values (e.g., if a student never accessed an assignment page) were replaced with zeros.

2.3 Course Clustering

As discussed in Section 1.3, there is no single, universally-valid relationship between student activity and student engagement; different courses will have different norms for what constitutes engaged student activity. These differences should be reflected by differences in how students behave between courses. Consequently, differences and similarities in aggregated course tool use can identify common usage patterns. To account for and understand these differences, we used a conventional clustering approach at the course level, using typical student behaviors (median values for each feature) to classify the courses. First we calculated, for each course site and each feature, the median of all enrolled students' feature values. We then used k-means (with 25 random start values) to determine cluster membership for each course from the unscaled median values, across all 19 feature dimensions. Exploratory analyses found that six (k = 6) clusters provided a visible inflection point in the incremental reduction of intracluster distances, and also provided suitable results, segregating separable patterns of student activity across courses without overfitting to esoteric courses. The centroids of each cluster's features are shown in Table 1; our code for determining these clusters is publicly available at https://osf.io/7cbzx/.

2.4 Engagement Ratings

Our institution routinely encourages all instructors to report their observations of student engagement across nine dimensions: Attendance; Participation; Assignments; Overall Engagement; Area of Concern; Quizzes and Exams; Writing Skills; Quantitative Skills; and Leadership. The first four categories reflect general behavioural properties of student engagement in a course, while the latter five categories measure specific skills and performance. Instructors register their ratings for any enrolled student on a restricted-access website (called the *Student Engagement Roster*) in any of these categories at any time during the academic term, and instructors can also include recommendations for improvement.

Ratings are immediately sent to the student, and some ratings also trigger personalized follow-up from a Success Coach or Academic Advisor, depending on the campus. Instructors are told to submit "observations" and "feedback" regarding student engagement in the provided categories, and that this is "critical to student engagement, learning, achievement, persistence and graduation."

All past courses where an instructor submitted ratings in any category for any student were eligible for the current analysis (see Section 2.1, above). However, considering the study's focus on student behaviours and not skillset or performance (which may be influenced by many factors external to the course), we limit the scope of our definition of engagement to ratings in the first four categories: Assignments, Attendance, Participation, and Overall Engagement. Any student who received a negative engagement rating in these categories (*Completing some but not all assignments, Not completing assignments, Never attended, Irregular attendance, Stopped attending, Inconsistent participation, Low participation, Sudden decline in engagement, or Not passing course*) was considered to have received a negative engagement indicator for the current analysis.

2.5 Modeling

To investigate the relationship between features of student activity and instructors' ratings of student engagement, we fit 6 mixed-effects logistic models (one for each cluster). For each model, the output variable was 0 or 1 (corresponding to the absence or presence of a negative engagement indicator, respectively), the fixed-effects were the activity features (z-scored within each course; see Table 2), and the random-effect was the course. Models were tested for collinearity, and fixed-effects were removed if the variance inflation factor (VIF) was greater than 6. For the small number of cases when a student had more than one enrollment in a single cluster, those observations were down-weighted so that each unique student contributed the same weight to the model for that cluster. Models fits were obtained with maximum likelihood estimation, using the lme4 package for R [34]. For some analyses described below, we converted the estimated values (log odds) to binary predictor values by determining the optimal cutoff in order to minimize misclassification errors, separately within each of the 6 models. Our threshold for determining statistical significance is whether the 95% confidence interval of an estimated effect does not include 0. Our code for this modeling work is publicly available at https://osf.io/a85en/.

3 RESULTS AND DISCUSSION

3.1 The Relationship between Activity Logs and Instructor Ratings of Engagement

Out of the 11,926 student enrollments in the current analysis, 1,324 received some form of negative engagement indicator from the instructor (11.1%), with 1,213 unique students receiving such flags. Each student enrollment was eligible to receive multiple ratings from different engagement categories, and on average, instructors provided negative ratings in 1.47 engagement categories (out 4) for enrollments receiving at least one flag. The 1,324 flagged enrollments included 704 negative ratings regarding Assignments (53.1%), 585 negative ratings regarding Attendance (44.2%), 301 negative ratings regarding Participation (22.7%), and 360 negative ratings regarding Overall Engagement (27.2%). In our logistic models, enrollments receiving any such rating were considered to have received a negative engagement indicator.

Model performance, as well as coefficient values and their associated 95% confidence intervals for each student activity feature, are shown in Table 2. Combining the estimates from all 6 clusters, estimators derived from LMS activity features explained 42.9% of the variance in instructors' negative engagement indicators. When converting model estimates to binary values (with cutoff thresholds selected to minimize misclassification error), the models correctly predicted 644 of the 1,324 instructors' flags (48.6%). Full model summaries are publicly available at https://osf.io/54dtj/.

As expected, the models for different clusters of courses differed somewhat in the quality of their relationships with instructor ratings of engagement. This variability provides additional support for previous researchers' claims that the relationship between student activity and student outcomes will be mediated by instructional context. Most notably, courses in Cluster 2 had no significant relationship between the features of LMS activity logs and instructor ratings of engagement. It should be unsurprising that, in a large sample of real courses, a small number of observations will deviate from norms; and indeed, Cluster 2 was the smallest cluster in our sample (6 courses; 224 unique students), with the lowest incidence of negative engagement indicators (3% receiving flags)[1]. But it *is* surprising, however, that while Cluster 2 made relatively sparse use of LMS tools, it did not represent the *lowest* level of tool usage. Indeed, Clusters 1 and 5 both had lower values than Cluster 2 for number of submissions, number of visits (sessions), duration of sessions, and page views per session, and even so, the relationship between activity logs and instructors' ratings of engagement in the models for Clusters 1 and 5 were both significant (explaining 54 and 33% of the variance in instructor ratings, respectively). Thus our results indicate that the relationship between activity logs and student engagement is not moderated merely by the *amount* of tool usage in LMS, but also by the *form* of tool usage. Relatively small amounts of LMS activity can be highly diagnostic of student engagement, so long as this activity is germane to the instructor's norms for student behavior within the context of the course.

With the exception of the model for Cluster 2, the remaining models accounted for between 33% and 66% of the variance in instructors' ratings of student engagement. For each of these, the estimators derived exclusively from features of LMS activity logs

[1] The six courses in Cluster 2 were a 200-level Information Systems course, a 300-level Journalism careers course, a 200-level Criminal Justice course, a 100-level Chemistry course, a 200-level Information Technology course, and a 300-level Greek History Course, distributed across 4 campuses.

were significantly related to the presence of negative engagement flags as raised by the instructor.

Across these models, the feature that is most diagnostic of instructor ratings of engagement is the total number of submissions. Again, excepting the model for Cluster 2, all other models had significant negative coefficients for the number of submissions, meaning that the more submissions students recorded in the LMS, the *less* likely they were to receive a negative engagement indicator. Another notably large effect, specifically in the models for Clusters 4 and 5, also had significant negative coefficients for the number of web sessions (which might be considered the number of unique visits to the course site), suggesting that more visits decreases the likelihood of negative engagement indictors, although these effects were marginal for Clusters 1, 3, and 6.

It makes sense that the number of submissions recorded in the activity logs should have this strong inverse relationship with negative engagement indicators. After all, engagement ratings related to Assignments (*Completing some but not all assignments, Not completing assignments*) made up the majority of flagged enrollments. However, even when removing these Assignment-based ratings from our criteria for what defines a negative engagement indicator, the number of submissions still remains a significant predictor of negative engagement in Clusters 1, 4, 5, and 6, and full models for all clusters (except Cluster 2) are still statistically significant under this depleted definition of engagement.

Other features did not bear such a consistent relationship between the amount of activity and engagement ratings. For example, in Cluster 1, viewing assignment pages *after* the deadline was marginally associated with better engagement (lower likelihood to receive a negative engagement indicator), while in the remaining clusters, the relationship was the opposite: viewing assignment pages after the deadline increased the odds of receiving negative engagement flags. Presumably for courses in Cluster 1, the process of reflecting on past assignments is characteristic of an engaged learner, but in Clusters 3-6, viewing assignments after the deadline indicated deficiencies in engagement. This kind of inconsistent relationship between student activity and student engagement was also evident in the duration of the longest period of inactivity within the course site, where clusters 1 and 3 had negative coefficients (longer periods indicated *decreased* likelihood to receive negative indicators), while clusters 4 and 6

Features of Student Activity (z-scored by course)	**Estimated Coefficients and 95% Confidence Intervals for Logistic Models corresponding to each Cluster** 1	2	3	4	5	6
Time on asgmt pages (m)						
Avg time between first access & asgmt deadline (h)						
Avg session duration with asgmt views (h)						
Avg page views / session with asgmt views (c)						
Visits to 'Files' after an asgmt view (c)						
Visits to other 'Assignments' after an asgmt view (c)						
Visits to 'Modules' after an asgmt view (c)						
Visits to 'Pages' after an asgmt view (c)						
Total asgmt views with no subsequent visit (c)						
Visits to other Canvas tools after an asgmt view (c)						
Number of asgmt submissions 6am-6pm (c)						
Number of asgmt submissions 6pm-midnight (c)						
Number of asgmt submissions midnight-6am (c)						
Total number of submissions (c)						
Total visits to asgmt pages before deadline (c)						
Total visits to asgmt pages after deadline (c)						
Number of unique sessions with site visits (c)						
Visits to Canvas's 'Calendar' of assignments (c)						
Longest period of inactivity within the site (h)						
r^2	0.54	0.18	0.66	0.44	0.33	0.63
(contrast with empty model) X^2	69.9	4.2*	50.0	538.2	248.8	156.4

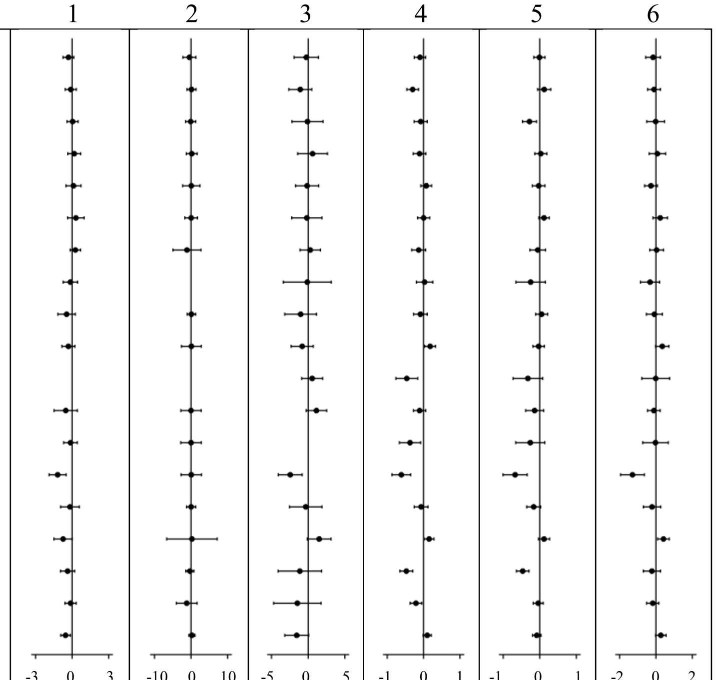

Table 2: Relationship between features of student activity and instructor ratings of engagement. Note: 'asgmt' is short for assignment; (m) indicates minutes, (h) indicates hours, (c) indicates count. Points represent the estimated coefficient value, and error bars show the surrounding 95% confidence interval. *Different models have different coefficient scales*. Negative coefficient values indicate that increases in that feature are associated with decreased likelihood to receive a negative engagement indicator. Model for cluster 2 is rank deficient for "Visits to 'Pages' after an asgmt view," and this feature is excluded from the model. 'Number of asgmt submissions' variables dropped from models for clusters 1, 2, and 3 to correct multicollinearity. * *not significant*.

had marginally positive coefficients (longer periods indicated *increased* likelihood to receive negative indicators). The inconsistency in coefficient signs for this feature is theoretically attractive because Conjin et al. [9] also found this factor to be diagnostic of course performance across the majority of courses, even while the signs of model coefficients for this factor (positive or negative) were inconsistent across their sample. It would seem that for some kinds of courses, being judicious about one's access of online resources (with potentially longer periods of inactivity) is a positive indicator of student engagement.

Importantly, we make no claims about the quality of these models of student engagement, nor boasts about our models' performance. On the contrary, we imagine that it would be easy to improve predictive performance by expanding our feature list with a wider range of relevant student actions in the LMS, or by testing alternative models (e.g., SVM or neural networks). The goal of this article is not to propose an optimal method of classifying engagement from activity logs (which may not exist), but rather to use multiple regression as a means of combining features of student activity in a transparent way, to ultimately assess the existence of relationships between activity logs and student engagement.

We also make no causal or directional claims regarding the relationship between engagement and student activity. It is entirely possible that, when registering their ratings in the engagement roster, instructors are using LMS records to determine which students get flags. If this is the case, the fact that the instructor (who has the final authority on deciding what constitutes engagement within a course) uses LMS activity records as a proxy for engagement ratings would provide a strong endorsement of the construct validity of LMS activity data for measuring engagement.

In our analysis, we find that there are positive relationships between features of students' interactivity within the LMS and with instructors' subjective ratings of their students' levels of engagement. This is a boon for learning analytics, because it validates the use of analytical models derived from activity logs as a diagnostic tool for the automatic detection of students who are disengaged. However, we also observe that there is not a single generic relationship between activity and engagement, and that what constitutes behavioral components of "engagement" will differ between courses. Our current approach involves accounting for these differences by classifying courses according to normative tool usage (the median value of all activity features from enrolled students) and then classifying students according to the students' individual deviations from the course norm (*z*-scores of individual features).

Having established these estimators of student engagement directly from Canvas web logs, we extend our analysis to the next logical step in assessing a practical implementation of this diagnostic tool, by investigating whether the modeled estimator is predictive of student outcomes in a way that is comparable to instructors' ratings of engagement.

3.2 Estimated Engagement Scores, Instructor Ratings, and their Relationships with Grade Outcomes

To evaluate the relationship between student engagement and course outcomes, we adopted a common metric that combines instances when a student receives a D or F in a course, or withdraws (sometimes abbreviated as DFW for institutional benchmarking). All 11,926 enrollments in the current sample had a single assigned grade (course sites with multiple or ungraded enrollments were excluded), and 2,157 of these grades were a D, F, or W (18.1%), which are defined as negative grade outcomes for the current analysis.

Table 3 shows the frequency of enrollments receiving these negative grade outcomes, split by whether the instructor levied a negative engagement rating on the student, and the model's estimate of negative engagement (derived from activity logs and converted to binary values as described above).

Table 3: Frequency Table of Engagement and Negative Grade Outcomes.

Instructors' Negative Engagement Ratings	Model Estimates of Negative Engagement	Negative Grade Outcomes 0	1
0	0	9,031	1,344
0	1	159	68
1	0	330	314
1	1	247	433

Of the enrollments that received a negative engagement rating from the *instructor*, 56.4% ultimately received a negative grade outcome. Moreover, of the enrollments that the *model* estimated to have low engagement, 55.2% ultimately received a negative grade outcome. Overall (combining all clusters), the interaction between instructor ratings and model estimates for predicting grade outcomes was significant, $z = -2.625$, as determined by a logit model with mixed effects for cluster and course, indicating that instructor ratings were reliably more sensitive than model estimates for the full sample. The sensitivity (d') of detecting negative grade outcomes for instructor ratings and model estimates in individual clusters are shown in Table 4.

Of course, if one were interested to fit the activity models to predict grade outcomes, there is little doubt that sensitivity of model estimates for this response variable could improve significantly. After all, the recent literature is replete with examples of researchers successfully predicting course performance from LMS activity [4-14]. However, the incremental gains in predictive performance if one were to predict grade outcomes, and not engagement, likely stem from the dispositional properties of successful students [9], rather than from the variance accounted for by student engagement as estimated in the present work.

Table 4: Comparison of sensitivity (*d'*) for negative grade outcomes from engagement ratings. Higher values indicate improved detection of students who have negative grade outcomes.

Cluster	Instructors' Negative Engagement Ratings	Model Estimates of Negative Engagement
1	1.73	1.42
2	1.81	1.12
3	1.32	1.41
4	1.18	1.22
5	1.30	1.01
6	1.41	1.39
Total	1.27	1.17

What is noteworthy about the present analysis, however, is that instructor engagement ratings are not *unconditionally* more sensitive than the model estimators of engagement for predicting course performance across all clusters. In Cluster 4 for example, which had the largest number of students, instructor estimates were slightly *less* sensitive than model estimates at predicting negative grade outcomes. We therefore infer that, when estimating a student's level of engagement, modeled indices from activity logs are not consistently missing some essential piece of variance that is globally fundamental for successfully detecting at-risk students. We heartily acknowledge that in many cases, instructors will have superior insights into the behaviors of their students and how these behaviors relate to the standards of achievement in their courses. It is not particularly remarkable that instructors' ratings have a statistically-significant advantage over our simple models at identifying students who are at risk. However, in some (easily classifiable) cases, our analysis also demonstrates that student activity within the LMS will reveal features of engagement that may outperform instructors' subjective assessments.

How then, can model estimates be deployed at scale to identify and support disengaged students, while still acknowledging that instructors have privileged insights in some contexts? Rather than circumventing instructor ratings, we imagine that model estimates might be displayed to the instructor in an LMS-based dashboard, and instructors could also be given the opportunity to endorse modeled indicators or make changes as necessary. Such a system might capitalize on the automated detection of disengaged students made possible by LMS activity logs at scale, while still augmenting model estimates with manual flags. Moreover, the situations where instructors override model estimates would provide valuable training data to iteratively improve classification performance.

4 CONCLUSION

The current study finds that features of student activity within the LMS can provide valid and useful indicators of student engagement. We observe that, at a large scale, estimators derived from LMS activity logs explain a significant proportion of variance in instructors' subjective ratings of student engagement, and moreover, this variance attributable to engagement also provides a usefully sensitive measure at identifying negative grade outcomes. However, we also join past researchers in asserting that there is no one-size-fits-all definition of engagement, as the structure of our modeled engagement indicators varied between different kinds of courses. Activity indices also varied in the strength of their correlations with instructor ratings across courses – for a category containing 1.3% of courses (Cluster 2), the correlation was not significant – so the validity of using activity logs to assess student engagement will be contingent on whether the analysis is sensitive to structural differences between courses. The current study offers one such method of quantifying these differences, by clustering courses according to students' normative levels of tool usage.

The purpose of this study was to test the construct validity of activity logs as a measure of student engagement. We were motivated by the strong momentum in the broader learning analytics community to predict student performance directly from this LMS activity data, and we saw it necessary to confirm that these activity logs were measuring what we assumed they were measuring about student behavior. As reviewed in the introduction, it would have also been possible for LMS activity to reflect deficiencies in student engagement, and if this were the case, the utility of these predictors would have been moot. The practical utility of a risk estimate is not merely whether it accurately classifies students according to their future outcomes, but also whether it reveals something actionable and remediable.

Our focus on actionable and remediable patterns of behavior guided the current operationalization of student engagement. We used instructors' ratings of students' levels of engagement along specific behavioral dimensions (as submitted to an institutional student support system). This narrow definition has strong advantages: instructor flags reflect the unique standards of normative student behavior within the context of a specific course, as well as the unique threshold an instructor might set for detecting deviation on these dimensions; but by no means do these ratings provide a comprehensive assessment of student engagement. To implement a complete assessment of student engagement would require students' self-reports of their own academic goals and experiences [32], affective and physiological measurements [28], and much more [30] – ultimately this effort would produce an index of engagement so diffuse as to obscure any practical use of the data.

Understanding and predicting student engagement within any educational context through an indicator — behavioral or otherwise – present in activity logs represents a means, not an end, of learning analytics research and applications. This assumption should be kept in mind in the development of all predictive models for student learning and performance. Indeed, the application and use of these models should be considered in relation to how they enable more formative and effective learning and pedagogical practices in addition to the overall model performance [35, 36]. Consequently, we view this work as establishing foundations for future applications of learning analytics to positively transform the practice of educators and student support infrastructures.

In this frame, the development of models for predicting students' behavioral engagement can be viewed as support for instructors to reflect on both their and their students' practices within the objects and expected norms of their courses. Further, the development of interventions based on automated detection systems to support teaching and learning require larger considerations in how such systems are intended to mediate and transform the practices under consideration. Simply introducing a system that predicts performance from activity logs, for example, without insight into how it is anticipated to be used in practice is akin to finding a problem for a solution to solve. Identifying the elements and points where such tools are theoretically meaningful and functionally useful across the multiple levels of educational activity must be considered in the further refinement and development of analytic models.

Looking forward, we hope the current validation study encourages learning analytics researchers to develop design-based [37] analyses of the applications of activity logs and experimental [38] analyses of their benefits in context. Through this process we hope the current work supports the development of effective and ethical analytic solutions that enable the positive transformative promise of applying data informed insights to educational practice.

ACKNOWLEDGMENTS

The authors thank Matthew Jones at University of Michigan, Stefano Fiorino, Jason Kaetzel, and Tori Scott at IU, and Greg Rommel at Instructure for their assistance as we developed and validated our techniques for extracting activity features from Canvas web logs.

REFERENCES

[1] J. Kim, R. LaRose, and W. Peng. 2009. Loneliness as the cause and effect of problematic internet use: The relationship between internet use and psychological well-being. *CyberPsychology & Behavior,* 12, 4, 451-455. DOI: https://doi.org/10.1089/cpb.2008.0327

[2] D. Clow. 2012. The learning analytics cycle: closing the loop effectively. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 134-138. New York, NY. DOI: https://doi.org/10.1145/2330601.2330636

[3] J. Pomerantz and D. C. Brooks. 2017. *ECAR Study of Faculty and Information Technology*. ECAR, Louisville, CO.

[4] A. Cicchinelli, E. Veas, A. Pardo, V. Pammer-Schindler, A. Fessl, C. Barreiros, and S. Lindstädt. 2018. Finding traces of self-regulated learning in activity streams. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, Sydney, Australia. DOI: https://doi.org/10.1145/3170358.3170381

[5] R. Cerezo, M. Sánchez-Santillán, M. P. Paule-Ruiz and J. C. Núñez. 2016. Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education,* 96, 42-54. DOI: https://doi.org/10.1016/j.compedu.2016.02.006

[6] F. Okubo, T. Yamashita, A. Shimada, and H. Ogata. 2017. A neural network approach for students' performance prediction. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, Vancouver, British Columbia. DOI: https://doi.org/10.1145/3027385.3029479

[7] J. W. You. 2016. Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education,* 29, 23-30. DOI: https://doi.org/10.1016/j.iheduc.2015.11.003

[8] C. Brooks, C. Thompson, and S. Teasley. 2015. Who you are or what you do: Comparing the predictive power of demographics vs. activity patterns in massive open online courses (MOOCs). In *Proceedings of the Second ACM Conference on Learning @ Scale*, New York, NY. DOI: https://doi.org/10.1145/2724660.2728668

[9] R. Conjin, C. Snijders, A. Kleingeld, and U. Matzat. 2017. Predicting student performance from LMS data: A comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies,* 10, 1, 17-29. DOI: https://doi.org/10.1109/TLT.2016.2616312

[10] L. V. Morris, C. Finnegan and S.-S. Wu. 2005. Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education,* 8, 3, 221-231. DOI: https://doi.org/10.1016/j.iheduc.2005.06.009

[11] T. Yu and I.-H. Jo. 2014. Educational technology approach toward learning analytics: Relationship between student online behavior and learning performance in higher education. In *Proceedings of the Fourth International Conference on Learning Analytics And Knowledge*, Indianapolis, Indiana. DOI: https://doi.org/10.1145/2567574.2567594

[12] S. Joksimović, D. Gašević, T. M. Loughin, V. Kovanović, and M. Hatala. 2015. Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education,* 87, 204-217. DOI: https://doi.org/10.1016/j.compedu.2015.07.002

[13] A. Hernández-García, E. Acquilla-Natale, J. Chaparro-Peláez, and M. A. Conde. 2018. Predicting teamwork group assessment using log data-based learning analytics. *Computers in Human Behavior*. DOI: https://doi.org/10.1016/j.chb.2018.07.016

[14] R. Umer. 2017. On predicting academic performance with process mining in learning analytics. *Journal of Research in Innovative Teaching & Learning,* 10, 2, 160-176. DOI: https://doi.org/10.1108/JRIT-09-2017-0022

[15] K. S. Na and Z. Tasir. 2017. Identifying at-risk students in online learning by analysing learning behaviour: A systematic review. In *2017 IEEE Conference on Big Data and Analytics (ICBDA)*, Kuching, Malaysia. DOI:

https://doi.org/10.1109/ICBDAA.2017.8284117

[16] J. Milne, L. M. Jeffrey, G. Suddaby, and A. Higgins. 2012. Early identification of students at risk of failing. In *ascilite 2012: Future Challenges, Sustainable Futures*, Wellington, New Zealand, 2012.

[17] Y.-H. Hu, C.-L. Lo, and S.-P. Shih. 2014. Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior,* 36, 469-478. DOI: https://doi.org/10.1016/j.chb.2014.04.002

[18] R. Conjin, A. Van den Beemt, and P. Cuijpers. 2018. Predicting student performance in a blended MOOC. *Journal of Computer Assisted Learning,* 34, 5, 615-628. DOI: https://doi.org/10.1111/jcal.12270

[19] H. A. P. L. Perera and P. R. M. R. Perera. 2017. Evaluation of Online Learning Activities of Accountancy Students Based on LMS Logs. In *8th International Conference on Business & Information ICBI*, University of Kelaniya, Sri Lanka. DOI: http://repository.kln.ac.lk/handle/123456789/18437

[20] Y. Park, J. H. Yu, and I. H. Jo. 2016. Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education,* 29, 1-11. DOI: https://doi.org/10.1016/j.iheduc.2015.11.001

[21] S. Dang, M. Yudelson, and K. Koedinger. 2017. Detecting diligence with online behaviors on intelligent tutoring systems. In *Proceedings of the Fourth ACM Conference on Learning @ Scale*, Cambridge, MA. DOI: https://doi.org/10.1145/3051457.3051470

[22] S. M. Jayaprakash, E. W. Moody, E. J. M. Lauría, J. R. Regan, and J. D. Baron. 2014. Early alert of academically at-risk students: An open source analytics initiative. *Journal of Learning Analytics,* 1, 1, 6-47. DOI: https://doi.org/10.18608/jla.2014.11.3

[23] J. R. Eyink, B. A. Motz, G. Heltzel, and T. M. Liddell. 2018. Self-regulated studying behavior, and the social norms that influence it. *PsyArXiv Preprint*. DOI: https://doi.org/10.31234/osf.io/bf6kg

[24] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic. 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. DOI: https://doi.org/10.1016/j.iheduc.2015.10.002

[25] S. J. Quaye and S. R. Harper. 2014. Student engagement in higher education: Theoretical perspectives and practical approaches for diverse populations, New York: Routledge.

[26] B. Motz, M. Rickert, T. Busey, and D. Landy, "Finding topics in enrollment data," In *Proceedings of the 11th International Conference on Educational Data Mining*, Buffalo, NY, 2018.

[27] A. Reschly and S. Christenson. 2012. Jingle, jangle, and conceptual haziness: Evolution and future directions of the engagement construct. In *Handbook of Research on Student Engagement*, Boston, Springer, 3-19.

[28] S. D'Mello, E. Dieterle, and A. Duckworth. 2017. Advanced, analytic, automated (AAA) measurement of engagement during learning. *Educational Psychologist,* 52, 2, 104-123. DOI: https://doi.org/10.1080/00461520.2017.1281747

[29] J. G. Greeno and Middle School Mathematics Project Group. 1998. The situativity of knowing, learning, and research. *American Psychologist,* 53, 1, 5-26. DOI: https://doi.org/10.1037/0003-066X.53.1.5

[30] G. M. Sinatra, B. C. Heddy, and D. Lombardi. 2015. The challenges of defining and measuring student engagement in science. *Educational Psychologist,* 50, 1, 1-13. DOI: https://doi.org/10.1080/00461520.2014.1002924

[31] A. Olney, E. F. Risko, S. K. D'Mello, and A. C. Graesser. 2015. Attention in educational contexts: The role of the learning task in guiding attention. In *The Handbook of Attention*, Cambridge, MA, MIT Press, 623-642.

[32] G. D. Kuh. 2009. The national survey of student engagement: Conceptual and empirical foundations. *New Directions for Institutional Research,* 141, 5-20. DOI: https://doi.org/10.1002/ir.283

[33] W. Lee and J. Reeve. 2012. Teachers' estimates of their students' motivation and engagement: being in synch with students. *Educational Psychology,* 32, 6, 727-747. DOI: https://doi.org/10.1080/01443410.2012.732385

[34] D. Bates, M. Maechler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software,* 67, 1, 1-48. DOI: https://doi.org/10.1080/01443410.2012.732385

[35] D. Gašević, S. Dawson, and G. Seimens. 2015. Let's not forget: Learning analytics are about learning. *Tech Trends,* 59, 1, 64-71. DOI: https://doi.org/10.1007/s11528-014-0822-x

[36] K. Kitto, S. B. Shum, and A. Gibson. 2018. Embracing imperfection in learning analytics. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, Sydney, Australia, 2018. DOI: https://doi.org/10.1145/3170358.3170413

[37] L. Lockyer, E. Heathcote, and S. Dawson. 2013. Informing pedagogical action: Aligning learning analytics with learning design. *American Behavioral Scientist,* 57, 10, 1439-1459. DOI: https://doi.org/10.1177/0002764213479367

[38] B. Motz, P. Carvalho, J. de Leeuw, and R. Goldstone. 2018. Embedding experiments: Staking causal inference in authentic educational contexts. *Journal of Learning Analytics,* 5, 2, 47-59. DOI: https://doi.org/10.18608/jla.2018.52.4