# The Positive Impact of Deliberate Writing Course Design on Student Learning Experience and Performance

Alia Lancaster[1], Scott Moses[2], Martyn Clark[3], Megan C. Masters[4]

**Abstract**

Learning management systems (LMSs) are ubiquitous components of the academic technology experience for learners across a wide variety of instructional contexts. Learners' interactions within an LMS are often contingent upon how instructors architect a module, course, or program of study. Patterns related to these learner interactions, often referred to as learning analytics implementation (LAI), can be represented by combining system-level LMS data with course-level design decisions to inform more granular insights into learner behaviour. The purpose of this paper is to use the LAI framework, specifically the principles of coordination and comparison (Wise & Vytasek, 2017), to examine how learner interaction patterns associated with LMS-use variables correspond to deliberate learning design decisions and course outcomes for a group of courses in the same undergraduate writing program. Visualizations of learner activity exhibited similar patterns of learner engagement across courses, corroborating the observation that design decisions heavily influence learner behaviour. Predictive analyses demonstrated strong influence of LMS use on final grades while accounting for course instructor. That is, while page views were not related to final grade, the length of discussion entries was often predictive. These results suggest that students who practised writing more — the main learning objective of this course — had higher final grades, regardless of variations in instructor and semester.

**Notes for Practice**

- Learning analytics implementation is a framework by which learning analytics metrics can be connected to learning design to best serve learners.

- In an undergraduate writing course, length of discussion entries, but not page views, predicts final grade.

- Differences among instructors can provide programmatic insights into the student experience of interacting with instructional content.

**Keywords**

Learning analytics implementation, learning management system, student engagement

Corresponding author [1]Email: abiller@umd.edu Address: 0117 Edward St. John Learning and Teaching Center, University of Maryland, College Park, 20742, USA, ORCID ID: https://orcid.org/0000-0001-7320-2349
[2]Email: pmoses1@umd.edu Address: Department of English, University of Maryland, College Park, 20742, USA, ORCID ID: https://orcid.org/0000-0001-7903-1456
[3]Email: mclark19@umd.edu Address: 0117 Edward St. John Learning and Teaching Center, University of Maryland, College Park, 20742, USA, ORCID ID: https://orcid.org/0000-0001-7452-7743
[4]Email: mkrol@umd.edu Address: 0117 Edward St. John Learning and Teaching Center, University of Maryland, College Park, 20742, USA, ORCID ID: https://orcid.org/0000-0003-4410-6141

## 1. Introduction

Learning analytics implementation (LAI), or the process of introducing the intentional and systematic use of learning-related, system-level data to inform input to the learning design and subsequent learner uptake processes, has received attention among researchers in recent years (Wise, Vytasek, Hausknecht, & Zhau, 2016; Wise & Vytasek, 2017). Previously constrained to more narrow analyses of learners' patterns of behaviour within online instructional platforms, LAI offers a more holistic approach by engaging with designers and instructors to discuss the patterns of learner behaviour expected to appear within a platform, given a specific course design. Put another way, LAI is "the purposeful framing of activity surrounding how analytic tools, data, and reports are taken up and used as part of an educational endeavor" (Wise & Vytasek, 2017, p. 152). This paper

explores the application of LAI principles within an authentic classroom context with the goal of determining how patterns associated with learning management system (LMS) usage relate to the intentional design and structure of a group of writing courses. Though the project reported here did not initially set out to use the LAI framework, it provides a useful way to conceptualize and organize the results of this programmatic learning analytics (LA) project and can serve as an example of how to operationalize this relatively new theoretical framework.

The LAI approach to course design triangulates three essential perspectives into a given learning context: (1) instructional goals and objectives, (2) learning design conceptual frameworks and methodologies, and (3) student experience within an instructional course. Building on these foundational considerations of LAI, McKee (2017) asserts that having access to LA is not the same as wanting to use LA or knowing how to effectively integrate it into teaching or administration practices. That is, while the goal of LA research is often to visualize and understand behavioural patterns across large populations of learners, the interpretation of and resulting decisions made based on these data are typically confined to the parameters surrounding an individual course context. Addressing this disconnect, Wise and Vytasek (2017) introduce the principles of coordination, comparison, and customization as a way of bridging the gap between learning design activities (the pedagogical course requirements associated with a given course) and LA (the resulting system-level data representing learners' behaviours when engaging with pedagogical content). As defined by Wise and Vytasek (2017, p. 153), the principle of coordination involves active collaboration between LA researchers and instructors to determine, prior to course implementation, where expected patterns in LA data would be found given a specific learning design. The principle of comparison involves identifying appropriate reference points, or benchmarks, against which a set of observed LA-related data can be compared. Lastly, the principle of customization involves the recognition and validity of the differential needs of the variety of audiences that consume LA-related data. All stakeholders, including database engineers, data scientists, instructors, learners, and program evaluators, require customized access to LA-related data to inform the context in which they work and the decisions made based on the data provided.

The current work mostly addresses the principles of coordination and comparison, with the principle of customization being the next natural step. The overall goal of the project is to provide actionable and fair programmatic evaluation metrics to online sections of a writing course taken by all juniors and seniors at a large R1 institution (a doctoral university with very high research activity). While instructors must have autonomy to structure their class and connect with students using methods that are individualized and appropriate (i.e., pedagogical intent), it is important that the writing program ensure that all students enjoy learning experiences that similarly contribute to their success in the course no matter the instructor. To that end, a course shell within the institution's LMS was created and populated with assignments, modules, discussions, and more by the second author and is made available for instructors teaching online sections of this course.

The goals of the analysis were twofold: (1) to validate the design of the course shell as meeting the best practices for online writing curriculum (i.e., applying the principle of coordination) and (2) to determine useful LMS metrics given natural variations that will occur among instructors (i.e., applying the principle of comparison). One group of potential end users for this effort is instructors; thus, they serve as references to one another for the principle of comparison. Within the program, there is much more variation among instructors than is represented in the current analysis (e.g., differences in delivery format); therefore, we hope that by statistically modelling the small differences that exist in our current dataset we can produce a method whereby any relative comparisons made are used programmatically to start conversations around certain pedagogical choices and potential impacts on the student experience.

First, we operationalize our research questions, informed by the course structure as designed by the second author and implemented with five instructors of the same course. While all instructors utilized the same online course shell with the same content, the pedagogical intent, or the degree to which instructors adapted or changed the content to best suit their and their students' needs, was also under investigation. Thus, we describe the educational context in which the work occurred. Using a heat map to visualize patterns in learner activity across course spaces, this study subsequently attempts to compare hypothesized patterns, given learning design decisions made by instructors, against observed trends in LMS system-level data and discuss the implications of converging or diverging patterns of learning activity. Next, working from the LA metrics, we empirically model observed variability across semesters and instructors in order to statistically determine the systematicity of course implementation practices. Finally, we examine which learning design decisions and LA metrics relate to learner performance and consider whether these metrics can serve as opportunities for early identification or intervention for learners themselves. By identifying differences in how instructors design their courses, and by comparing expected and observed patterns in observed LA data, we offer a methodology in which LA data can be effectively used to provide insight into the efficacy of learning design decisions. The future intent of the project is to successfully integrate LAI into the teaching and administration practices of a university writing program.

## 1.1. Research Questions

This paper investigates the following four research questions (RQs), which apply to the principles of coordination and

comparison when noted.

**RQ1:** Do instructor hypotheses concerning expected online patterns in learner engagement, such as page views, correspond to observed LA metrics pulled from an LMS given the intentional learning design of a given program (principle of coordination)?

**RQ2:** Does a combination of LA metrics predict student performance within undergraduate writing program courses (principle of coordination)?

**RQ3**: Do LA metrics offer opportunities for early detection of learners who tend to perform best in the course (principle of comparison)?

**RQ4:** Does pedagogical intent influence student performance (principle of comparison)?

Investigation into the four research questions outlined above builds empirically on the work of Wise and Vytasek (2017), which focuses on the intentional connection between LA metrics and LAI. In partnership with instructors, these analyses may help to determine which types of LA data are most relevant to instructors, students, and program administrators. Results may yield useful empirical insights into recommendations concerning the principle of customization, that is, which data are hypothesized to be most useful to which stakeholder audience and at what point within the administration of a given course.

## 2. Educational Context and Methods

### 2.1. Educational Context

This study focuses on an undergraduate writing program at a public, mid-Atlantic R1 university with more than 37,000 students. The program offers 14 courses that fulfill one of the university's undergraduate general education requirements. These courses include science writing, business writing, technical writing, and legal writing. Approximately 85 program instructors teach nearly 7,000 undergraduate students every year.

To teach such a large student population, courses are offered in multiple delivery formats, including traditional face to face, blended or hybrid, and asynchronous online. While a majority of instructors use the university's LMS, Canvas, only instructors teaching in asynchronous online formats are *required* to do so and receive training to meet general online course design standards informed by Quality Matters[1] and online writing instruction specific standards informed by the Conference on College Composition and Communication (CCCC)[2].

With instructors using a variety of delivery formats, with some choosing not to use the LMS at all, and some not having received the same training or having the same expectations of LMS usage as others, an effective LAI across the writing program cannot be one size fits all. Therefore, we are choosing to start by looking at a subset of the writing program. Specifically, we investigate five instructors teaching the same course, with the same asynchronous online delivery format, during both the fall 2018 and spring 2019 semesters. These five instructors each chose to use the same course shell, offering the same course modules, requiring the same major assignments, and maintaining the same schedule of deadlines. The exception is one instructor, who limited the number of discussion board activities, while the other four instructors used all discussion board activities. The selection of these five instructors is a convenience sample; having one instructor who chose not to use the course shell in its entirety was not part of an experimental or quasi-experimental design; rather, it was an insight gleaned from the initial data collection.

The course is structured as a writing-intensive experience designed to help students build the writing skills that they will likely need after graduation, focusing particularly on tailoring writing for specific audiences' background knowledge and expectations. The course shell design is intended to mimic a writing workshop and to position professional writing as a complex, iterative process. Course content is divided into eight learning modules. Goals for each of the modules are clearly visible in the course space. With the first half of the semester designed for skills development and the second half of the semester designed for skills application in the form of the final project, the course shell template drives more activity in the LMS during the first half of the semester.

In preparation for each major assignment, students read exemplar texts, participate in targeted discussions, produce drafts, and provide peer feedback to other students in the class. Discussion board activities embedded in the course shell represent the activity that students participate in most frequently throughout the semester. The required discussion board posts serve different purposes, fitting within three different categories: editing exercises, example review, and social interaction. As Morgan (2011)

---

[1] Quality Matters (QM) is a non-profit "organization leading quality assurance in online and innovative digital teaching and learning environments." Among a range of products, QM offers a set of Higher Education Course Design Rubric Standards available at https://www.qualitymatters.org/qa-resources/rubric-standards/higher-ed-rubric.

[2] In March 2013, the CCCC Committee for Effective Practices for Online Writing Instruction released "A Position Statement of Principles and Example Effective Practices for Online Writing Instruction (OWI)," found at https://cccc.ncte.org/cccc/resources/positions/owiprinciples.

and other researchers have noted, discussion boards are not homogeneous spaces, and they are not always used effectively. The variety of discussion board opportunities in this particular writing course correspond to many of the same categories that Ransdell, Borror, and Su (2018) outline in their list of effective uses of discussion boards, with the goal of creating active learning environments that motivate students to participate and connect with each other on an emotional level. While integral to the student experience of this writing course, discussion boards do not contribute significantly to students' final grade. In the course shell, traditional discussion board activities are just a portion of other formative online activities, cumulatively representing only 15% of the final grade in the course.

The course shell is designed to facilitate a collaborative learning experience, with students working together as peers as they progress through the modules (e.g., discussions and peer review). The final submissions of major assignments include reflections on the assignment and the peer review process in addition to the core target document. Requirements are clearly articulated in the course space, and the descriptions for the assignments include detailed information about the expected length of various components of each assignment, including expectations for discussion posts and information about the contribution of each component to the overall project or course grade. Leveraging the full functionality of the LMS, all of the content and intended learning processes are built into the course shell, which instructors can use with virtually no need for customization.

The data in this investigation can be considered multilevel in nature, with students nested under instructors and instructors nested under semester. Instructors themselves were not strictly nested but instead cross-classified under semester such that each instructor taught each term (Figure 1). All retrospective data mining efforts were approved by the institutional review board office.
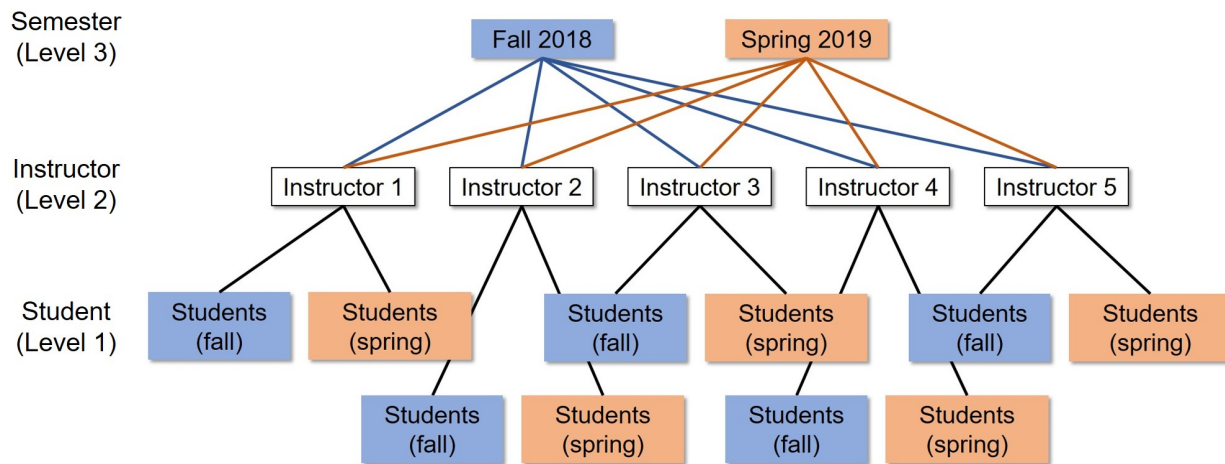


**Figure 1.** Multilevel analytical structure visual model

## 2.2. Data Mining Procedures

Data for 609 students in the target instructors' courses were extracted using rCanvas, an R API wrapper for the Canvas LMS API (Ranzolin, 2016). Note that the models described use only complete cases, meaning students who have observations for all variables, so the number of students in each model may differ slightly and is noted in the model descriptions. These data serve as student behaviour variables. Page views are the count of pages in the course that a user accesses, aggregated by day. Page views encompass a student's view of any course page and were not related to the final grade. The text of the discussion posts and replies was also extracted. Final grades were extracted from the LMS course gradebook as the overall percentage grade in the course. With the exception of the word count of the discussion entries, the page views metric and threaded discussions are available to all instructors through the standard LMS interface.

The two LMS variables also represent two different types of LA variables. Page views represent a variable that is available in most LMSs as well as other academic technologies and thus has the potential to be the most generalizable. We propose that length of discussion entries is a metric that is specifically valuable to these online writing courses because it represents both student engagement and development of a skill necessary to succeed in the course (i.e., writing). Thus, while not all courses incorporate discussions, they are fairly important in the realm of online writing instruction, and undergraduate writing instruction more broadly, which is a common feature in many higher education institutions in the United States (National Census of Writing, 2017).

## 2.3. Analysis Procedures

A series of visualizations and mixed-effects linear models was performed to examine the LMS behaviour variables and to unpack the relationship between the LMS behaviour variables and the course outcomes. The LMS behaviour variables were

transformed for the inferential statistical analysis described below due to large differences in scale; however, they are presented in the raw scale for descriptive and visualization purposes and for ease of interpretation by the reader.

To provide context to the inferential statistics and to validate the course design, we begin with descriptive statistics and visualizations. The heat maps in the following section, featuring student activity across the semester, are especially useful when determining whether student activity proceeds as intended. Our target course was designed to have overall more student activity in the beginning of the semester to facilitate the writing necessary for the final project due at the end of the semester. Therefore, we predicted that page views, as one measure of student activity within the online course space, would be greater for most students in the first half than in the second half of the semester. The descriptive statistics address the first research question and the principle of coordination, in that the heat maps reveal whether predicted and designed usage patterns exist as intended by the course design. There is also an element of the principle of comparison when comparing the average final grade, page views, and length of discussion entries among the instructors to determine what degree of variation actually exists. Although not predictive, we have found providing such descriptive statistics directly to LA users, instructors in this case, to be very powerful by making patterns visible that were previously invisible.

For the remaining three research questions, we used inferential statistics in the form of linear mixed-effects models. While a full description of mixed-effects models is beyond the scope of the current paper, we hope the following provides enough information to be transparent in our model parameter choices and to allow for replication. Summary and interpretations in the subsequent sections were written intentionally with a very broad and applied audience in mind. The linear mixed-effects models were conducted with the most recent version of the lme4 package (Bates, Maechler, Bolker, & Walker, 2016) in R (R Core Team, 2016). Mixed-effects models were chosen over other analysis types, such as regression and hierarchical linear models, due to their advantages for this dataset, including the ability to increase power, to incorporate variables that are not completely normally distributed, and to incorporate cross-classified rather than fully nested random effects (i.e., as in our data, where instructors are crossed rather than fully nested within semesters; Baayen, Davidson, & Bates, 2008; Linck & Cunnings, 2015). All models were run as forced-entry models for fixed effects. The structure of random effects differed per model and is discussed below.

The outcome variable was typically represented as final grade in the course, which had a possible range of 0% to 100% and was not transformed. All continuous predictor variables were transformed into a $z$-score. Thus, estimates for the continuous variables can be interpreted as the change in final grade percentage due to one standard deviation change in the continuous variable, for the average value of the other continuous variables and for the reference level of all categorical variables. All models included two categorical variables: *semester* and *instructor*. The reference level for *semester* was fall 2018, with the other level being spring 2019. *Instructor* contained five levels, one per instructor, with Instructor 1 being the reference level.

Contrast coding was applied to all mixed-effects models, and the significance of each fixed-effect term was tested using a Satterthwaite approximation for denominator degrees of freedom (Satterthwaite, 1946) and $F$-statistic using the afex R package (Singmann, 2019). Contrast coding allows for an ANOVA-like output, such as the main effect of condition collapsing across all other factor levels and for the sample-mean value of any continuous variables. The afex package provides $F$- and corresponding $p$-values for each parameter by eliminating each fixed effect from the model one by one. Therefore, an ANOVA table output for each model is provided that answers questions such as whether a main effect of semester on final grade was established. For full model outputs, please see the supplemental documentation in Appendix A.

**Table 1.** *N*-Size per Cluster

| Instructor | Number of Students | | |
| | Fall 2018 | Spring 2019 | Total |
|---|---|---|---|
| Instructor 1 | 75 | 37 | 112 |
| Instructor 2 | 71 | 73 | 144 |
| Instructor 3 | 37 | 19 | 56 |
| Instructor 4 | 74 | 74 | 148 |
| Instructor 5 | 75 | 74 | 149 |

For pairwise comparisons of fixed effects, estimated marginal means were utilized to understand the direction and specific comparisons in the significant main effects and interactions. The R package emmeans (Lenth, 2018) was employed, which allows for a Tukey correction method for multiple comparisons. For example, if the main effect of *instructor* was significant on final grade, the estimated marginal means for each instructor were generated to compare final grades among all instructors, across the other conditions, and for the sample-mean value of the continuous variables. Table 1 displays the *N*-sizes per cluster. Because the data contain fewer than the recommended 30 clusters, we utilized corrective techniques where possible. A

Kenward–Roger adjustment was applied to estimate the fixed-effect standard errors, and a restricted maximum likelihood estimation (REML) approach was taken to avoid biasing estimates of variance components (McNeish & Stapleton, 2016). In addition, we aimed to produce a parsimonious model while maintaining an accurate representation of the multilevel nature of the data.

## 3. Results

### 3.1. Descriptive Statistics (Principle of Coordination)

In order to address the first research question and the principle of coordination, Figures 2 and 3 provide visual representations of student activity over the entire course as heat maps. Time is represented along the *x*-axis, and each student is represented along the *y*-axis, with a single row per student. Activity, in terms of the number of page views for a particular student on a particular day, is represented by a coloured rectangle. A blank rectangle indicates that a student did not have any page views during a particular day. The reader should be able to discern from the dips in page views where breaks occurred in the fall and spring semesters. Darker rectangles indicate more activity than lighter rectangles. A student with activity every day would appear as a solid horizontal bar of varying colour. These heat maps are similar to, although distinct from, the student "waterfall" visualizations by Wiley (2011) and Forteza, Whitmer, Fritz, and Green (2018). Both mapped student activity per week, instead of per day as we do, and student activity was operationalized as time in the LMS course space for the former and all possible student activity in the LMS for the latter.
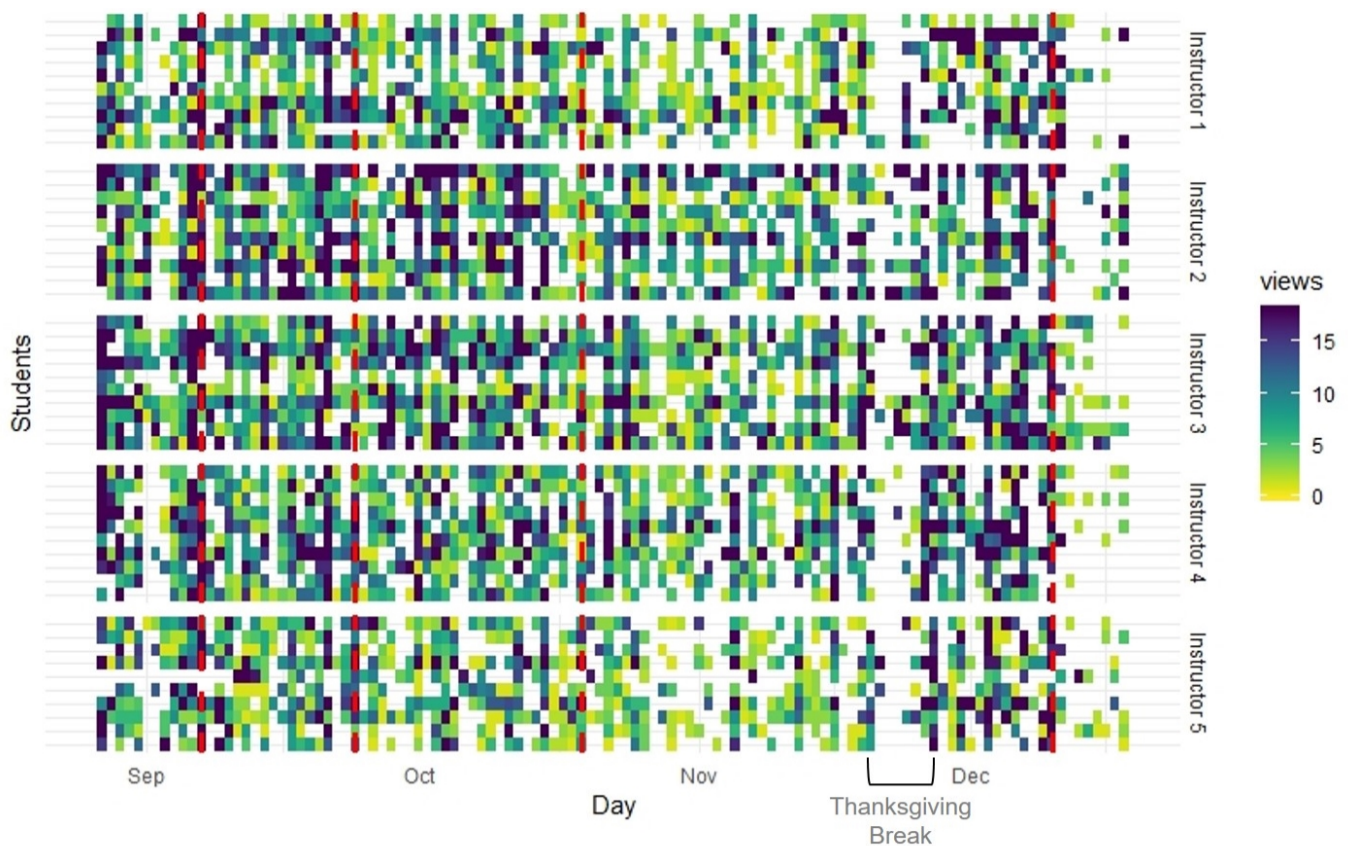


**Figure 2.** Example fall 2018 course activity for 10 random students per instructor

To provide an easier comparison for the reader, the sample_n function from the dplyr package (Wickham, François, Henry, & Müller, 2018) was used to randomly select 10 students per instructor per semester, and the page views for those students are represented in these figures. (Complete versions of these visualizations with all students are available in the supplemental documentation in Appendix A.) The colour scale was based on the activity of all students across all instructors in a given semester. Because outliers with extreme numbers of page views can mask the differences between more typical students when using this type of colour gradient, the upper limit of the colour scale was capped at the 80% quantile for each semester. Thus,

the darkest colour indicates 18 or more page views for fall 2018 (Figure 2) and 26 or more page views for spring 2019 (Figure 3).
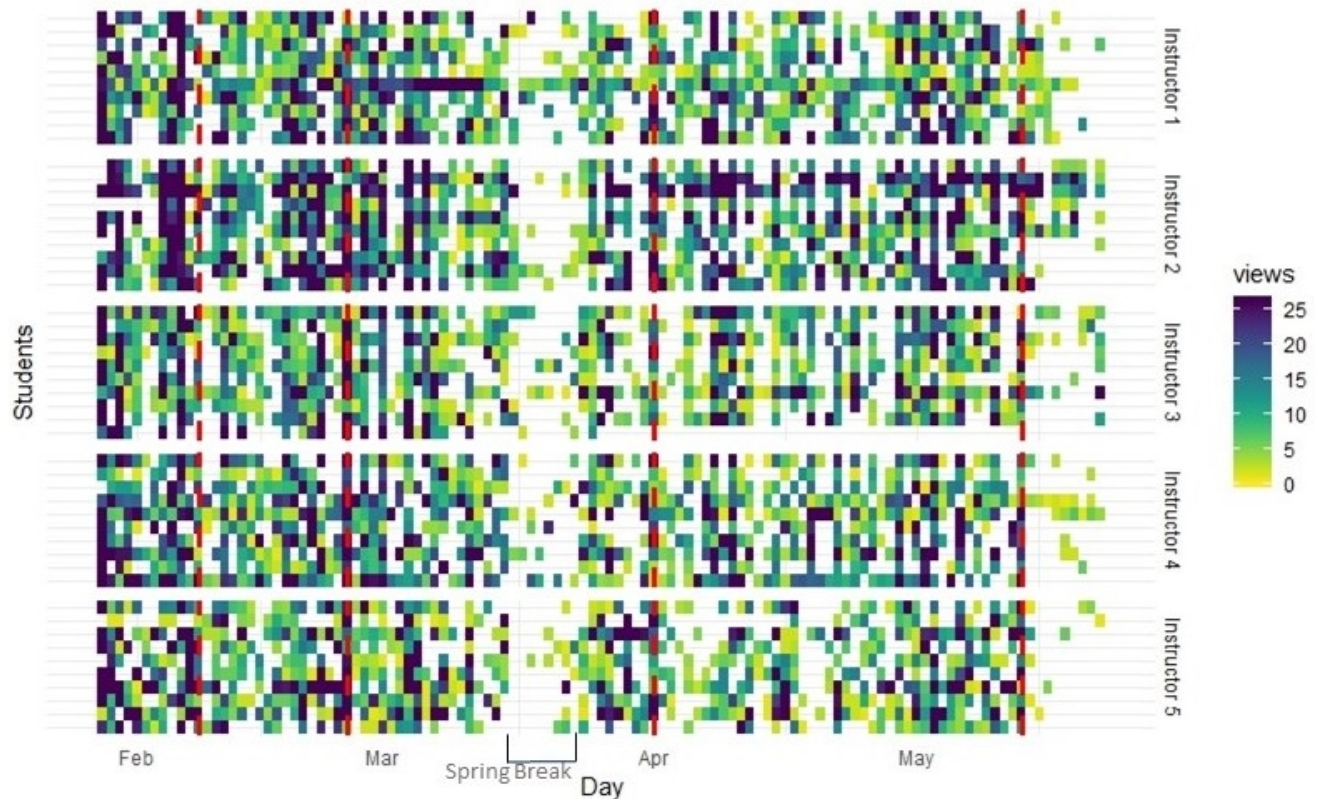


**Figure 3.** Example spring 2019 course activity for 10 random students per instructor

By visually inspecting the heat maps, one can see that even in the same course taught by the same instructor, reviewing the activity patterns of a random subset of students reveals a variety of activity patterns. None of the students were active in the LMS course space every day of the semester, though some showed fewer days of inactivity than others. As mentioned, the course shell design depends on relatively more assignments due during the first half of the semester than the second, since these activities are considered necessary prerequisites for successfully completing the final project. The heat maps generally reflect this trend, in that activity tends to wane toward the middle of the semester. Dashed vertical lines in the figures indicate key points in the semester, including the end of the add/drop period and the submission deadline for each of the three major assignments.

Therefore, with the LA metric page views, it is possible to observe the intentional design of the course, which ties into the principle of coordination. In other words, Figure 3 visually confirms the intentional front-loading of the course, focused on skill development (i.e., darker shades representing more page views), with later course content focused on skill application (i.e., lighter shades representing fewer page views), with more work likely happening outside of the LMS. Similar to the student waterfall activity visualizations in Forteza et al. (2018), only page views are necessary to capture the intention of the course shell. Therefore, this visual analysis encompasses the principle of coordination (specifically that of conceptual coordination) in that the hypothesized learning actions were observed in the page views (i.e., student activity) in the online course space.

To understand the variation among instructors, Figure 4 displays the average final grade per semester and instructor. Mean final grade is represented by the large, opaque dot, and the smaller, transparent dots represent observations of each student's final grade, so that the spread is visible. Semester average of final grade across instructors is indicated with a black, vertical line. The five instructors are represented by different colours and on different rows. Instructor 2 is slightly below both semester averages, while Instructors 1, 4, and 5 are slightly above both semester averages. Instructor 3 is above average in the fall and below average in the spring.
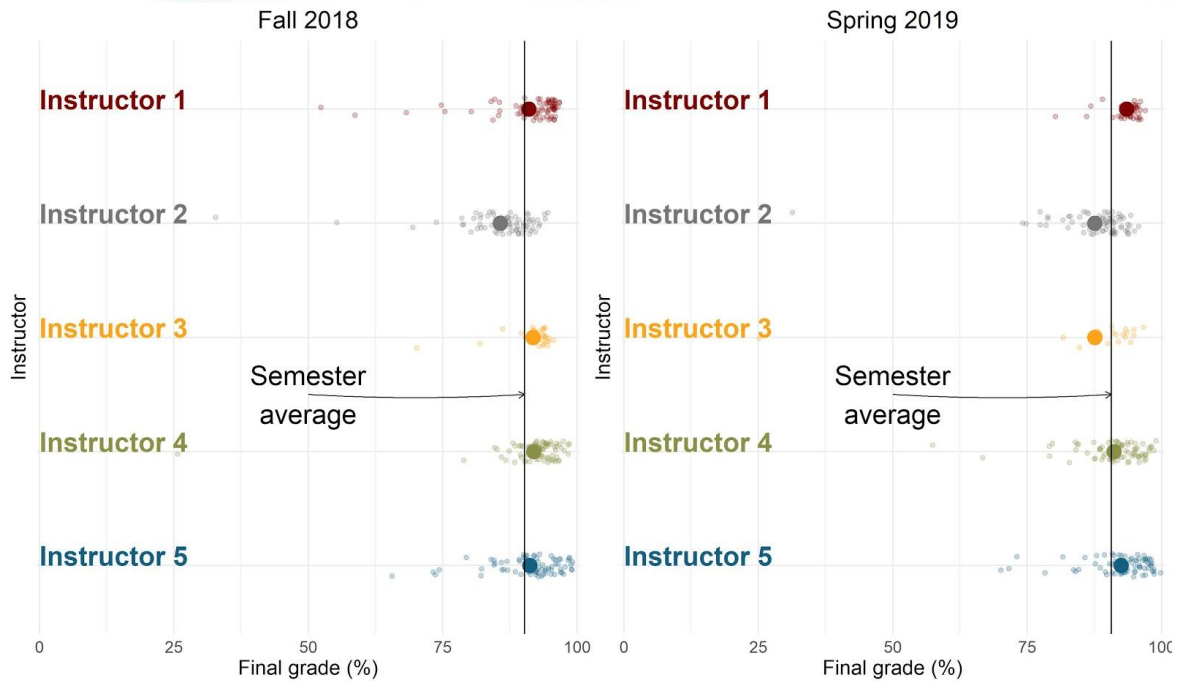
**Figure 4.** Final grade average (large dot) and spread (smaller dots) per student, instructor, and semester

The subsequent figures display the two predictors *page views* and *word count of discussion board entries* in a similar manner. Figure 5 depicts page views, and students under Instructor 5 have fewer cumulative page views on average, which likely relates to the smaller number of required discussions for that instructor. Figure 6 displays cumulative length of discussion entries per student, and it is clear that students under Instructor 5 wrote less on discussion entries than students under the other instructors, who used the course shell in its entirety. Although the first research question is addressed by the heat maps, the principle of comparison also comes into play, as the intended future users of these descriptive analytics are instructors and program administrators. Thus, being able to see the average values and spread of not only final grades but page views and length of discussion entries allows instructors to calibrate themselves against one another and program managers to have a high-level overview of the course outcomes and activities.
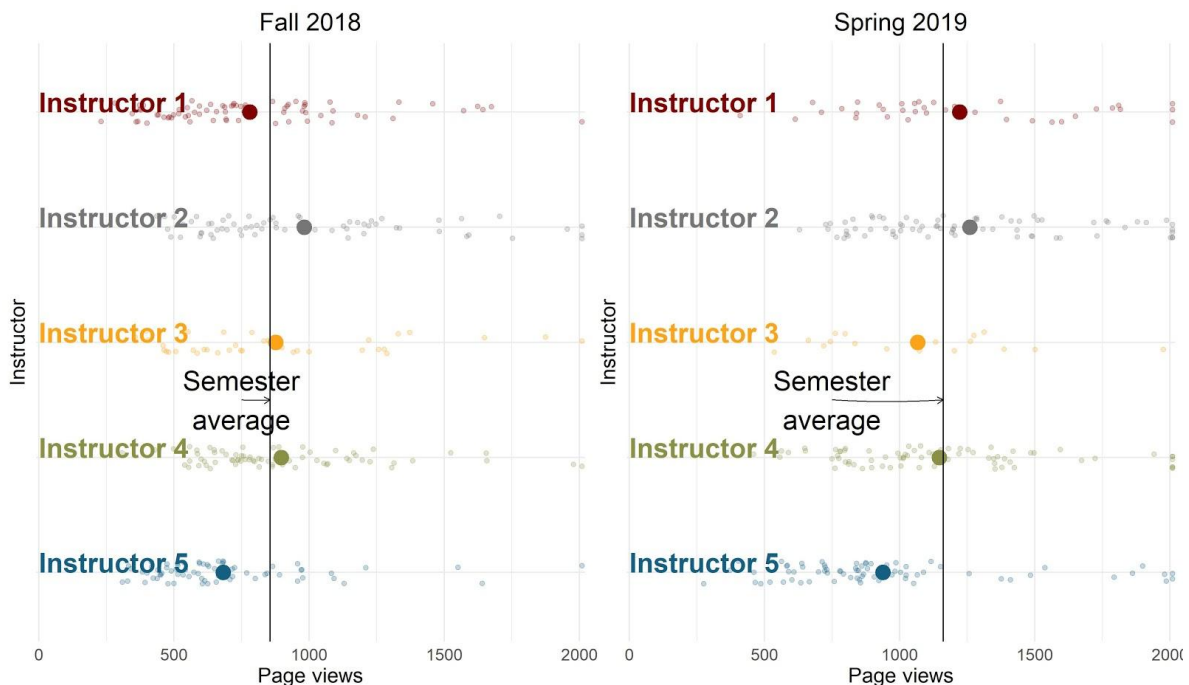


**Figure 5.** Cumulative page views average (large dot) and spread (smaller dots) per student, instructor, and semester
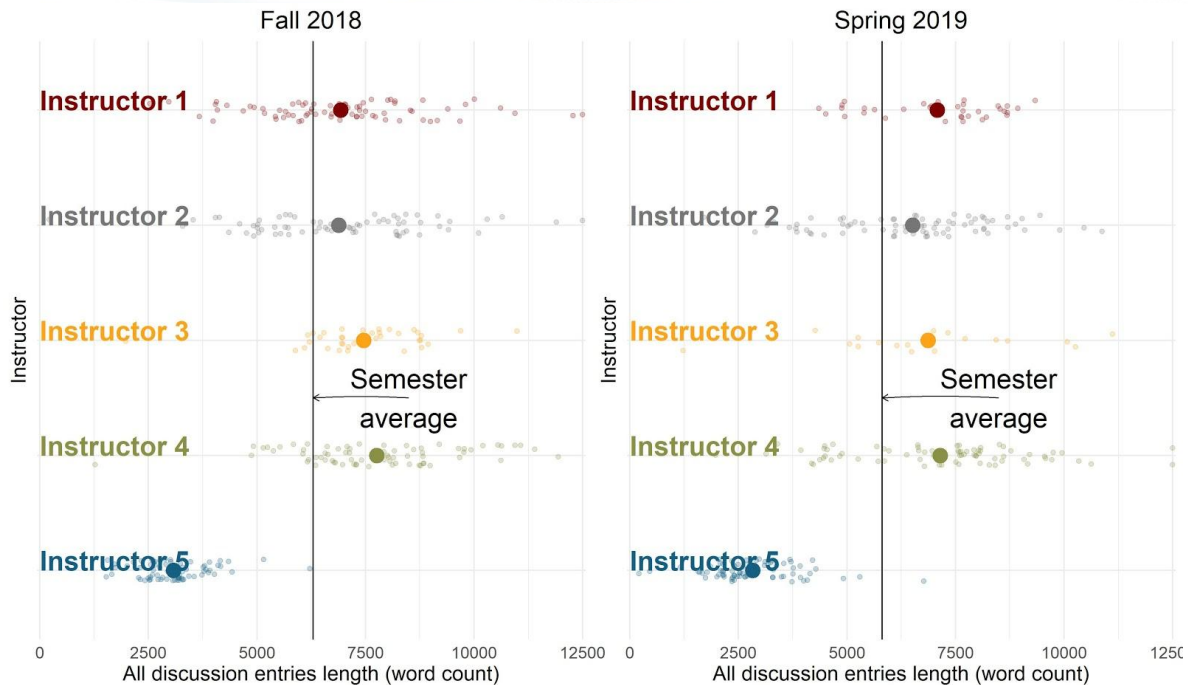
**Figure 6.** Cumulative discussion entries in word count average (large dots)
and spread (smaller dots) per student, instructor, and semester

Table 2 details the mean, standard deviation, and range of each predictor, untransformed, used in the following multilevel models, across instructors and semesters. Note that all students passed the course and had at least some page views and discussion entries. Unlike other courses, including massive open online courses (MOOCs), which often appear in the LA literature, sometimes with extreme attrition, this writing course at this institution historically has a high pass rate. Note that both page views and length of discussion entries during the first two weeks have ranges that include zero or near-zero minimums, which were not unexpected due to smaller delineations of these two variables. For instance, it is possible that during the first two weeks a student who had not yet registered for the course did not write any discussion entries, resulting in a length of discussion entries of zero during that time. Moreover, although discussion posts did contribute to 15% of the final grade (see Section 2, Educational Context and Methods, for more details), some students did not post very many discussion entries (i.e., the range was 1 to 48 posts). It is also worth noting again that the grade points accrued from these entries were based solely on participation, not length, which is one reason we chose length of discussion entries rather than number of posts as the variable to examine in the quantitative analyses. Thus, in terms of the principle of coordination, some expected variation among instructors is visible in the descriptive statistics.

**Table 2.** Descriptive Statistics for All Variables

| Variable | Time Period | Mean | SD | Range |
|---|---|---|---|---|
| Final grade (%) | semester | 90.38 | 7.92 | 74.68–99.83 |
| Page views | semester | 995.02 | 518.69 | 231–5,607 |
| Length of discussion entries (words) | semester | 6,077.33 | 2,506.32 | 213–15,881 |
| Discussion posts count | semester | 34.81 | 11.24 | 1–48 |
| Page views | first two weeks | 178.70 | 116.00 | 3–1,471 |
| Length of discussion entries (words) | first two weeks | 1,070.86 | 441.39 | 0–2,706 |

### 3.2. Quantitative Analysis

The analyses in the following section address each of the research questions in turn. In Section 3.2.1, we investigate the effects of *instructor* and *semester* separately on final grade, page views, and length of discussion entries to address the first research question and to determine whether any of the patterns observed in the descriptive statistics remain or change once the nested nature of the data has been accounted for statistically. The analyses in Section 3.2.2 explore the effect of all of the variables together on final grade, which allows for the relative importance of each to appear as well as for the exploration of any

interactions. The analyses in Sections 3.2.2 and 3.2.3 are structurally similar, but in Section 3.2.3 the LMS behaviour is limited to just that exhibited during the first two weeks of the course. This truncation addresses the third research question about early intervention. The fourth research question about the effect of instructor variation is addressed and summarized throughout the sections.

### 3.2.1. Influence of Instructor and Semester on Grade and Student LMS Behaviour (Principle of Coordination)

The first set of mixed-effects models was able to directly determine the influence of instructors and semester on final grade, page views, and length of discussion entries. The purpose of these models was to ascertain whether some of the differences observable in the descriptive statistics are significant when accounting for the structured nature of the data (e.g., instructors nested within semesters) and to inform the selection of fixed-effect interactions in subsequent models. That is, did the observed variations among instructors and across semesters play a role in this course (i.e., the principle of coordination) and, if so, how did the instructors differ (i.e., the principle of comparison)?

Therefore, in three separate models, final grade, page views, and length of discussion entries, each acted as the outcome with fixed-effect predictors of the categorical variables *semester* (reference level fall 2018) and *instructor* (reference level Instructor 1). By including instructor as a categorical variable, we hope to capture some of the variations inherent in having multiple instructors of the same course who may have slightly different goals (i.e., pedagogical intent). Some of these differences are known and/or measurable reliably (e.g., the number of discussions), and some are not (e.g., personality). Random slopes and intercepts were included for instructors nested within semesters.

With final grade as the outcome, there was no significant effect of *semester* ($F(1, 4.37) = 0.19$, $p = 0.680$) and a marginal effect of *instructor* ($F(4, 3.89) = 4.74$, $p = 0.080$) on final grade percentage for the 609 students who had observations for all of these variables (i.e., complete cases), as determined by eliminating the fixed effects one by one and comparing models with and without them (i.e., model comparison was assessed with the $F$- and $p$-values provided). Because *instructor* only marginally affected the final grade, the instructor pairwise comparisons do not show any significant differences between any two instructors. The model-estimated final grades for each instructor ranged from 92.1% to 86.6%. This range is narrower than the purely descriptive range in Table 2 because the model-estimated final grade takes into account the random and fixed effects (i.e., the multilevel nature of the data).

In a similar model, page views were considered as an outcome variable (i.e., the same predictors, instructor, and semester, and the same random effects structure), and *instructor* was a significant fixed effect ($F(4, 3.74) = 7.78$, $p = 0.040$), while *semester* was not ($F(1, 3701.11) = 1.50$, $p = 0.220$), for the 591 students with complete cases. Model-estimated page views, as a $z$-score, indicated that students under Instructor 2 had 0.68 of a standard deviation more page views than students under Instructor 5, regardless of the semester ($t = 5.63$, $p = 0.049$).

With an outcome of length of discussion entries in words and the same predictors and random effects structure, there was a significant effect of *instructor* ($F(4, 3.71) = 156.06$, $p < 0.001$) but not of *semester* for the 609 students with complete cases. Model-estimated pairwise comparisons demonstrate that students under Instructor 5 wrote less than students under all other instructors (Instructor 5 compared to Instructor 1: $t = 17.69$, $p = 0.001$; Instructor 2: $t = 18.38$, $p = 0.003$; Instructor 3: $t = 15.33$, $p < 0.001$; Instructor 4: $t = 22.37$, $p = 0.002$).

Thus far, once the hierarchical nature of the data is accounted for, the semester of instruction does not significantly impact student final grade, page views, or length of discussion entries. However, there are differences among the instructors in terms of cumulative number of page views (i.e., students in Instructor 2's course had more page views than those in Instructor 5's course) and length of discussion entries (i.e., students in Instructor 5's course wrote fewer words in their discussion entries than students in other instructors' courses). Due to this known influence of instructors on page views and length of discussion entries, our subsequent models include not only the level 1 student LMS behaviour variables but also interactions of *instructor* with *page views* and *instructor* with *length of discussion entries*. Using relative comparison among the intended end users (i.e., instructors) from the principle of comparison, there do appear to be some meaningful differences among instructors with regard to influences on student behaviour (i.e., page views and length of discussion entries). The following analyses are meant to determine whether these differences, taking into account other known variables, impact student outcomes.

### 3.2.2. Final Grade Predictors during the Semester (Principles of Coordination and Comparison)

The following model inspects the relative effects of instructor, page views, and length of discussion entries on final grade. Specifically, the model included the effect of page views, length of discussion entries in words, and semester (i.e., the fixed effects) on final grade (i.e., the outcome). Note that all student-level LMS behaviour variables were transformed into $z$-scores and were relative to one another under the principle of comparison. We first attempted to include random slopes and intercepts for *instructor* nested within *semester*; however, there were convergence issues. Thus, we only retained random slopes and intercepts for *instructor*, removed *instructor* as a fixed effect, and retained *semester* as a fixed-effect covariate. It should be noted that models with and without the nested random effects structures produced similar results in terms of significance and effect size, and we are presenting the results of the most parsimonious models. Additionally, we included two fixed-effects

cross-level interactions: (1) length of discussion entries (i.e., level 1 — students) by instructor (i.e., level 2) and (2) page views (i.e., level 1) by instructor (i.e., level 2). These two interactions were included because *instructor* was a significant predictor of both length of discussion entries and page views in Section 3.2.1.

As seen in Table 3, there was a significant main effect of length of discussion entries and a significant interaction between instructor and length of discussion entries for the 591 students with complete cases. For the main effect of length of discussion entries, students who wrote longer discussion entries had higher grades ($b = 5.85$, $SE = 0.46$, $df = 551.05$, $t = 12.62$, $p < 0.001$) for the reference levels of Instructor 1 in the fall term and for the average number of page views. The beta weight suggests that for every standard deviation increase in length of discussion entries (i.e., 2,506 words as detailed in Table 2, which represents about the length of Sections 1.0 to 2.2 of the current paper), a student's final score rose 5.85% for the reference levels of Instructor 1 in the fall term and for the average number of page views.

**Table 3.** Final-Grade Mixed-Effects Model with Semester,
Page Views, and Length of Discussion Entries Fixed Effects

| Effect | df | F |
|---|---|---|
| Page views | 1,575.63 | 0.49 |
| Discussion entry length | 1,576.60 | 14.95*** |
| Semester | 1,575.33 | 2.71 |
| Discussion entry length × Instructor | 4,567.88 | 8.51*** |
| Page views × Instructor | 4,575.43 | 0.58 |

*Note.* ***$p < 0.001$. **$p < 0.01$. *$p < 0.05$. †$p < 0.10$.

While the interpretation of the main effect is useful, it is perhaps more accurate to consider the model-estimated slopes of length of discussion entries on final grade for each instructor, since the interaction was significant. Figure 7 presents these slopes, with each student's length of discussion entries in words along the *x*-axis, their model-predicted final grade along the *y*-axis, and the colour indicating the instructor. A steeper slope indicates a stronger effect of discussions on final grade. The slopes also have the advantage of being estimated across semesters and for the average number of page views. For the interaction, the basic pattern still held; for all instructors, there was a positive relationship between final grade and number of discussion entries. The exact statistical influence of instructor on this relationship (i.e., the interaction) can be assessed by comparing the slopes of the effect of length of discussion entries between instructors in a pairwise manner.
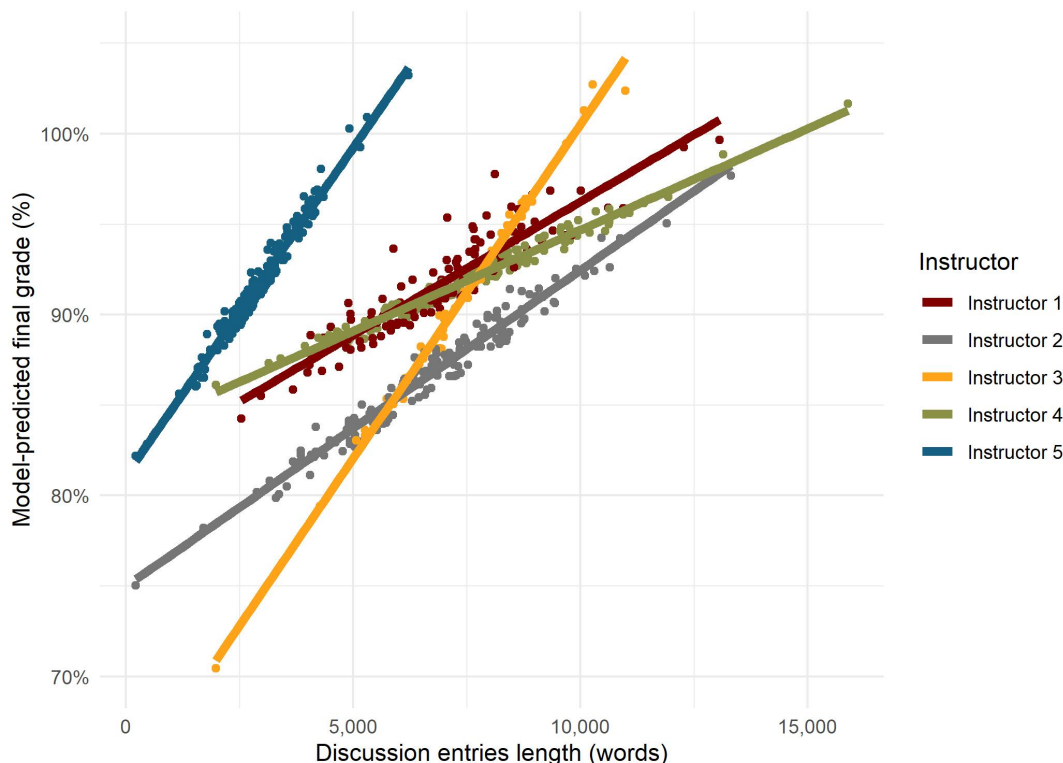


**Figure 7.** Model-estimated slopes of length of discussion entries on final grade per instructor

For instance, in Figure 7, every one standard deviation change in length of discussion entries led to a final grade increase of 3.34% for students under Instructor 1, across semesters and for the average page views. When these slopes are compared inferentially, the effect is even stronger (i.e., a significantly steeper slope estimate) for Instructor 3 than for Instructor 1 ($t = -4.12$, $p < 0.001$), Instructor 2 ($t = -3.71$, $p = 0.007$), and Instructor 4 ($t = -4.74$, $p < 0.001$). The effect was also stronger for Instructor 5 than for Instructor 1 ($t = -3.37$, $p = 0.007$), Instructor 2 ($t = -2.94$, $p = 0.028$), and Instructor 4 ($t = -3.83$, $p = 0.001$). Therefore, students who posted longer discussion entries tended to have higher final grades, and even more so for Instructors 3 and 5. In other words, the slopes for Instructors 3 and 5 in Figure 7 are significantly steeper than for the other instructors.

In addition to length of discussion entries, we also ran a similar model with count of discussion entries (mean = 34.81, SD = 11.24, range = 1–48) substituted for the variable of length of discussion entries, which yielded very similar results. Count of discussion entries is a metric that would be roughly accessible to an instructor throughout the semester, but we wanted to test a more sophisticated measure of discussion entries, namely, length. There are of course even more ways to operationalize a discussion entry (e.g., valence, number of ideas, type-token ratio), but this first pass may tell us whether it is worth creating a new metric for instructors that is not readily available to them in the LMS interface.

With regard to the second research question, the correlation of LA metrics to student performance, it appears that, at least for this online course taught by these five instructors, page views throughout the semester do not play a role in predicting final grade. Both the number of discussion posts and the slightly more nuanced metric of length of discussion entries do positively correlate with final grade, suggesting that writing practice does lead to better final grades in this writing course. These results are consistent with the principle of coordination. That is, while it is necessary to view course pages in this online course, discussion entries have a clear and explicit tie to the learning goals of the course and are therefore more related to learning outcomes. In terms of pedagogical intent and the fourth research question, Instructor 5 chose to have fewer online discussions, and the positive relationship between length of discussion entries and grade was even stronger for that instructor's students, suggesting that with fewer opportunities, the practice of writing in the discussion entries becomes more important. Students under Instructor 3 also displayed this stronger relationship, although the reason may be different. From the descriptive statistics, it appears that a few students under Instructor 3 had low numbers of posts and therefore the shortest cumulative length of posts in the sample.

### 3.2.3. Final Grade Prediction during the First Two Weeks (Principle of Coordination)

The third research question investigated whether the same metrics previously explored are possible metrics for future early interventions. In the course shell design, there was a clear and deliberate front-loading of the class, such that there were more assignments and online activities during the first half of the semester, which was reflected in more page views during that same time period, regardless of instructor. At this institution, the first two weeks of the semester are known as the add/drop period, during which students can add or drop classes without financial penalty. It is often the time when students who are on a waitlist are able to finalize their schedules. We chose to explore this time period because it is hopefully early enough that an intervention could be successful and because it has a clear end date each semester that may make it easier for instructors to act on. Additionally, there is an assumption that the nature of the course design was playing a role in driving activity in the LMS even at this early stage, as we saw in the heat map.

With this context, we ran a mixed-effects model with final grade as the outcome and semester as a categorical fixed effect. The change in this model was that the LMS behaviours of page views and length of discussion entries in words were summed over only the first two weeks of the semester. The analysis of interest was whether online course activity during the first two weeks of a course related to final grade. We also included interactions of page views by instructor as well as length of discussion entries by instructor.

**Table 4.** Final Grade Mixed-Effects Model with Semester, Page Views, and Length of Discussion Entry Fixed Effects Reflecting Activity during the First Two Weeks of the Semester

| Effect | df | F |
|---|---|---|
| Page views first 2 weeks | 1,589.24 | 1.77 |
| Discussion entries length first two weeks | 1,590.82 | 0.95 |
| Semester | 1,587.70 | 0.47 |
| Discussion entries length first 2 weeks × Instructor | 4,525.61 | 2.30† |
| Page views first 2 weeks × Instructor | 4,588.62 | 1.08 |

*Note*: ***$p < 0.001$. **$p < 0.01$. *$p < 0.05$. †$p < 0.10$

Table 4 depicts the significance of each fixed effect in this model for the 603 students with complete cases. There was a marginal and weak interaction of instructor and length of discussion entries, and the model-estimated slopes of length of discussion entries on final grade for each instructor revealed that the length of discussion entries had an even stronger effect on students under Instructor 3 than under Instructor 1 ($t = -2.86$, $p = 0.05$). By using the model-estimated slopes for this interaction, we learned that a one standard deviation increase in length of discussion entries during the first two weeks (i.e., 441.39 words) led to a 0.82% increase in final grade for students under Instructor 1 and a 3.43% increase in final grade for students under Instructor 5. Therefore, in terms of potential metrics for an early intervention, length of discussion entry, but not page views, might inform intervention or early alert system developments.

Overall, the inferential statistics yielded compelling results with respect to the use of LA metrics to inform course outcomes and the potential development of early warning systems. Regarding the second research question specifically, length of discussion entries but not page views rises to the top in terms of LA metrics that predict student performance in this online writing course. For research question three, length of discussion entries during just the first two weeks of the course do marginally predict final grading in an interaction with instructor, making it a candidate metric for any future early interventions. Lastly, those students who wrote longer discussion entries had higher final grades, and this relationship was often even stronger with the instructor who included fewer discussion entries, suggesting that developing the skills of writing, even when limited, does matter in a writing course.

## 4. Discussion

The impetus for the visualizations and analyses in the current paper was to determine the feasibility of using LA metrics from the institution's LMS in a program review framework, with instructors and program administrators being the final consumers of these metrics. The program in question was an upper-level professional writing program, the type of which serves thousands of learners across the United States (National Census of Writing, 2017). Therefore, the findings are applicable not only to the current institution but possibly to learners enrolled in similar writing programs and courses. To serve the large student population required to complete courses each year, courses are offered in a variety of formats. One format is taught completely online and utilizes a course shell that is pre-populated with course content, such as assignments and discussions. The current project had two goals: (1) to validate the design of the course shell and (2) to determine useful LMS metrics given natural variations that occur among instructors. These goals also offered an opportunity to operationalize the LAI framework (Wise & Vytasek, 2017) in a real-world setting, especially the principle of coordination (i.e., do LA metrics match the intended design of the course shell?) and the principle of comparison (i.e., how do instructors, as end users of these metrics, differ relative to one another?).

Instructional and academic freedom for instructors is important in the educational context of this writing program, but so, too, is ensuring a certain level of coherence of the educational experience across all students enrolled in the program. Using a course shell is not a requirement for online instructors, and instructors who choose to use the course shell are not restricted from making changes to it, such as adding or removing resources or activities. It is important to note that the course shell under investigation is not simply a repository of unorganized information; rather, it was designed to facilitate a specific collaborative and iterative learning process by leveraging LMS functionality, organized into modules, and rolled out over a semester. In this study's consideration of LAI, our main objective was to uncover how LA can help us visualize the intended design of this course shell and how, if at all, that design relates to learner performance. Moreover, after attempting to create a baseline of expected activity and performance, we consider how some deviations in the use of the course shell might then affect learner activity and performance. The four research questions operationalize our overall goal using specific LMS behaviour variables, page views and length of discussion entries, and address the principle of either coordination or comparison. The principle of coordination comes into play, for instance, when validating whether the course had more activity toward the beginning of the semester, as intended to allow students more time to work offline during the later half of the semester. The principle of comparison comes into play in the descriptive and statistical comparisons of the instructors to one another. Because instructors are the end users of these analytics, we hope the relative comparison of one instructor to the others is useful in illuminating how certain choices can influence student engagement and learning.

In our initial analysis of LMS page views (i.e., addressing the first research question), we found that student activity aligned generally with the intended learning trajectory envisioned by the course, with more student activity concentrated in the first half of the semester across all five instructors. Moreover, we found that on average students under Instructor 5 had both fewer page views and shorter discussion entries than students under the other four instructors, who used the course shell in its entirety. To consider the impact of this difference, we used page views and length of discussion entries as predictors of final grade (i.e., addressing the second research question), determining that length of discussion entries was indeed a meaningful predictor of learner performance. The more students wrote during the semester, the better they did in the course. Moving to the third research question, length of discussion entries even provided some predictive power during the first two weeks of the course,

thus serving as a possible metric for intervention, especially for a course designed to be more active in the beginning rather than the end of the term. In terms of pedagogical intent and research question four, on average, students under Instructor 5 wrote less, because they participated in fewer discussion board activities, making the relationship of length of discussion entries to performance even more impactful. Instructor differences like types and frequency of communication, which often happens outside the LMS, might further explain instructor-related differences among these otherwise identical courses. While it is beyond the scope of this paper to disentangle course shell modifications and other pedagogical differences that are unknown, we acknowledge that both of these as well as other features (e.g., instructor personality) could influence the role the instructor variable had in the multilevel models. That being said, *instructor* was not usually significant on final grade as a main effect, but only as an interaction with length of discussion entries, suggesting that course shell modifications were impacting student behaviour.

We did not find the page view variable to be a meaningful predictor among students enrolled in this online writing course. While the heat maps were useful in validating the course shell design and can be actionable, they were not predictive of final grade. We do not find these two phenomena contradictory; rather, it is possible that page views are useful for one setting but not another. The multilevel models point to the fact that while useful for course management and in course design, page views, at least in this program, are not indicative of student engagement in a way that directly impacts their final grade. There are a few possible reasons for this finding. One reason is that, while there is probably a minimum number of page views for more successful students (i.e., a student who never views course pages will probably not do well in an online class), there are a number of reasons for high page views. For instance, a relatively large number of page views could indicate a student who is doing well and who regularly reads the course material or engages with discussions with other students. It could also indicate a student who is struggling and who has not figured out the most efficient way to engage with online material. It is also possible that the predictive power of page views is dwarfed by that of length of discussion entries, which is much more directly related to the learning objectives of the course (i.e., writing). Even if page views had been predictive of final grade, our recommendation would not have been to require students to log in and view more pages, just as our recommendation for length of discussion entries is not simply to encourage students to write longer discussion posts. The recommendation from our findings is to use them a starting point to engage with instructors as to how their choices influence the student experience, soliciting their input on what they think a standard experience for a student in this course should be, regardless of instructor, and on how they think different writing opportunities throughout the course could reinforce the learning objectives.

The current study marks the initial attempt to incorporate LA intentionally using the LAI framework with a subset of courses housed within a larger undergraduate writing program. Vast research literatures compare online and face-to-face courses; passive learning and active learning environments; seemingly endless predictive metrics and models; and the specific learning benefits of LMS features, including discussion boards. In the case of this single undergraduate writing program, all of these issues, and more, exist at the same time, serving as a microcosm of the university at large and of other modern universities like it. Researchers like Fritz (2016) and Nguyen, Rienties, Toetenel, Ferguson, and Whitelock (2017) have already reached a level of scale allowing them to consider the relationships between design and performance across hundreds of courses and thousands of students. As these researchers can attest, reaching such a level comes after years of finding, accessing, and analyzing instructor, course, and learner data. Our study represents a single academic program, at a single university, near the beginning of this journey.

To begin thinking about applying LAI in the remaining courses within the undergraduate writing program or even university-wide, we must pose a range of questions: What do we know about the intended design of courses as they exist now? What do instructors teaching these courses intend for their students? What are the differences between these courses, and are these differences meaningful? And how can the LAI framework be leveraged to create positive curricular reform?

As noted by Cuban (1993), one of the reasons that curricular reform is so challenging is that there are actually multiple curricula at play in any given situation: what is officially mandated, what is taught by the instructor, what is learned by the student, and what is assessed. The course shell template used by these five undergraduate writing instructors was intended to embody the curricular goals of the undergraduate writing program. Even so, one of the instructors, Instructor 5, chose to limit the number of discussion activities used, and as a result of this decision, there was an even stronger relationship between discussion board activity and student performance. This stronger relationship suggests that fewer discussion board activities correlate with lower grades for some students in this course. Although this particular instructor may easily provide a pedagogically sound reason to modify the course shell, can providing LA metrics about discussion board activity from other course sections lead this instructor to reflect on the decision and potentially inform future versions of the course? Can such metrics help other instructors in the same program who are not using the course shell template reassess their decision?

The framework of Wise and Vytasek (2017) begins to address these questions. The principle of coordination allows for design decisions to be course specific. The designer of the course shell, and the online writing principles it embodies, was involved in this analysis, ensuring that the appropriate reference points were utilized. This practice can also inform metrics for other instructors with the same course who are not using the course shell. The next step is to involve more instructors and

program staff to obtain wider input now that a first step has been taken. For the principle of comparison, a relative reference frame was chosen such that student behaviour in the LMS was relative to other students under different instructors rather than using an absolute. Arriving at the principle of customization, how might these instructors access and use this information, or, instead, how might program administration use this information to create or change policies regarding online course design to ensure that the student experience is consistent across instructors and courses within the program? Armed with some initial LA insight, we must now determine what actions can be taken by instructors and program administrators to enhance future versions of the same course.

## 5. Conclusion

When systematically mined with intention and purpose, LMS data have the potential to make previously invisible patterns of student activity visible. Working with LA data, our goal in this paper was to provide an example of how the principles of coordination and comparison, as detailed in the LAI framework (Wise & Vytasek, 2017), relate to patterns associated with LMS use, learning design decisions, and course outcomes. Specifically, the analytics validated intentional design of a course shell used by five instructors in an online writing course, and the metric most closely related to the learning objectives, length of discussion entries, was most predictive of final grade (principle of coordination). Pedagogical intent, operationalized through instructors using more or fewer discussion boards, led to these opportunities to practise writing being even more impactful for students with fewer opportunities (principle of comparison). While this analysis only delved into five instructors, future work in this area may include developing programmatic metrics for instructors and students to ensure that, no matter the format, all students who enter the program have at least the same baseline experience.

## References

Baayen, H., Davidson, D., & Bates, D. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 309–412. https://dx.doi.org/10.1016/j.jml.2007.12.005

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2016). *lme4: Linear mixed-effects models using Eigen and S4* (Version 1.1-12). Retrieved from https://cran.r-project.org/web/packages/lme4/index.html

Cuban, L. (1993). The lure of curricular reform and its pitiful history. *The Phi Delta Kappan, 75*(2), 182–185. Retrieved from https://www.jstor.org/stable/20405055

Forteza, D., Whitmer, J., Fritz, J., & Green, D. (2018). *Improving student risk predictions: Assessing the impact of learning data sources*. Retrieved from https://www.blackboard.com/resources/improving-student-risk-predictions-assessing-impact-learning-data-sources

Fritz, J. (2016). LMS course design as learning analytics variable. In J. Greer, M. Molinaro, X. Ochoa, & T. McKay (Eds.). *Proceedings of the First Learning Analytics for Curriculum and Program Quality Improvement Workshop*, 25 April 2016, Edinburgh, UK (pp. 15–19). Retrieved from http://ceur-ws.org/Vol-1590/

Lenth, R. (2018). *emmeans: Estimated marginal means, aka least-squares means* (Version 1.2.2). Retrieved from https://github.com/rvlenth/emmeans

Linck, J. A., & Cunnings, I. (2015). The utility and application of mixed-effects models in second language research. *Language Learning*, *65*(S1), 185–207. https://dx.doi.org/10.1111/lang.12117

McKee, H. (2017). An instructor learning analytics implementation model. *Online Learning*, *21*(3), 87–102. https://dx.doi.org/10.24059/olj.v21i3.1230

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review, 28*(2), 295–314. https://dx.doi.org/10.1007/s10648-014-9287-x

Morgan, T. (2011). Online classroom or community-in-the-making? Instructor conceptualizations and teaching presence in international online contexts. *The Journal of Distance Education/Revue de l'éducation distance*, *25*(1). Retrieved from https://www.learntechlib.org/p/54018/

National Census of Writing. (2017). *2017 Four-Year Institution Survey*. Retrieved from https://writingcensus.swarthmore.edu/survey/4/year/2017

Nguyen, Q., Rienties, B., Toetenel, L., Ferguson, R., & Whitelock, D. (2017). Examining the designs of computer-based assessment and its impact on student engagement, satisfaction, and pass rates. *Computers in Human Behavior*, *76*, 703–714. https://dx.doi.org/10.1016/j.chb.2017.03.028

R Core Team. (2016). *R: A language environment for statistical computing* (Version 3.3.0). R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Ransdell, S., Borror, J., & Su, A. (2018). Users not watchers: Motivation and the use of discussion boards in online learning. *Florida Distance Learning Association Journal*, *3*(1), 4. Retrieved from https://nsuworks.nova.edu/fdla-journal/vol3/iss1/4

Ranzolin, D. (2016). *rCanvas: R client for Canvas API.* (Version 0.0.0.9001). Retrieved from https://github.com/daranzolin/rcanvas

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, *2*(6), 110–114. https://dx.doi.org/10.2307/3002019

Singmann, H. (2019). *afex: Analysis of factorial experiments*. (Version 0.27-2). Retrieved from https://cran.r-project.org/web/packages/afex/index.html

Wickham, H., François, R., Henry, H., & Müller, K. (2018). *dplyr: A grammar of data manipulation*. (Version 0.7.8). Retrieved from https://CRAN.R-project.org/package=dplyr

Wiley, D. (2011, February). *Openness, learning analytics, and continuous quality improvement* [Lecture slides]. Presented at the Educause Learning Initiative, 14–16 February 2011, Washington, DC, USA. Retrieved from https://www.slideshare.net/opencontent/openness-learning-analytics-and-continuous-quality-improvement-6975093

Wise, A. F., & Vytasek, J. M. (2017). Learning analytics implementation design. In C. Lang, G Siemens, A. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 151–160). SoLAR. https://dx.doi.org/10.18608/hla17.013

Wise, A. F., Vytasek, J. M., Hausknecht, S., & Zhao, Y. (2016). Developing learning analytics design knowledge in the "middle space": The student tuning model and align design framework for learning analytics use. *Online Learning*, *20*(2), 155–182. https://dx.doi.org/10.24059/olj.v20i2.783

## APPENDIX A: SUPPLEMENTAL DOCUMENTATION

This section contains supplemental information to accompany the following publication: Lancaster, A., Moses, S., Clark M., & Masters, M. (2020). The positive impact of deliberate writing course design on student learning experience and performance. *Journal of Learning Analytics, 7*(3), 48–63.

The first section contains tables for the full output of the multilevel models described in the main text, the second section contains sample code for running the statistical models, and the third section contains heat maps of page views for all students.

**Multilevel Model Tables**

The following two tables describe the sum coding applied to the categorical variables *instructor* and *semester*, respectively, to all models that contain those variables. Each column was entered as a fixed effect of the models, as indicated below with the noted reference level. Therefore, when the model parameters for "Instructor: Instructor 2" are presented below, the estimates apply to Instructor 2 compared to Instructor 1 for the reference level of the other categorical variables as well as for the overall mean of the *z*-scored continuous variables. For fixed effects, a colon indicates a categorical variable level (e.g., "Semester: Spring 2019"), and a "×" indicates an interaction (e.g., "Length of discussion entries × Instructor: Instructor 2"). For random effects, a "|" indicates a nesting relationship (e.g., "Instructor|Semester" means *instructor* nested within *semester*).

**Table 1.** Sum Coding Applied to Instructor Categorical Variable.
Reference Level Is Instructor 1

| Instructor | Instructor 2 | Instructor 3 | Instructor 4 | Instructor 5 |
|---|---|---|---|---|
| Instructor 1 | 0 | 0 | 0 | 0 |
| Instructor 2 | 1 | 0 | 0 | 0 |
| Instructor 3 | 0 | 1 | 0 | 0 |
| Instructor 4 | 0 | 0 | 1 | 0 |
| Instructor 5 | 0 | 0 | 0 | 1 |

**Table 2.** Sum Coding Applied to Semester Categorical Variable.
Reference Level Is Fall 2018

| Semester | Spring 2019 |
|---|---|
| Fall 2018 | 0 |
| Spring 2019 | 1 |

**Table 3.** Multilevel Model with Final Grade Outcome Predicted by Instructor and Semester

| Fixed effects | b | SE | df | t-value | p-value |
|---|---|---|---|---|---|
| (Intercept) | 90.44 | 0.51 | 3.14 | 177.20 | < 0.001 |
| Instructor: Instructor 2 | 1.64 | 0.98 | 2.63 | 1.68 | 0.205 |
| Instructor: Instructor 3 | −3.81 | 0.93 | 2.23 | −4.08 | 0.046 |
| Instructor: Instructor 4 | −0.28 | 1.13 | 4.53 | −0.25 | 0.816 |
| Instructor: Instructor 5 | 1.06 | 0.93 | 2.20 | 1.14 | 0.362 |
| Semester: Spring 2018 | −0.22 | 0.50 | 2.82 | −0.44 | 0.691 |
| **Random effects** | **Variance** | **SD** | | | |
| Instructor|Semester (Intercept) | 1.28 | 1.13 | | | |
| Term (Intercept) | 0.04 | 0.19 | | | |
| Residual | 58.23 | 7.63 | | | |

**Table 4.** Multilevel Model with Page Views Outcome Predicted by Instructor and Semester

| Fixed effects | b | SE | df | t-value | p-value |
|---|---|---|---|---|---|
| (Intercept) | 0.03 | 0.24 | 0.00 | 0.11 | 1.000 |
| Instructor: Instructor 2 | −0.01 | 0.09 | 2.77 | −0.14 | 0.898 |
| Instructor: Instructor 3 | 0.31 | 0.08 | 2.25 | 3.95 | 0.048 |
| Instructor: Instructor 4 | 0.00 | 0.11 | 7.13 | 0.03 | 0.980 |
| Instructor: Instructor 5 | 0.06 | 0.08 | 2.23 | 0.76 | 0.518 |
| Semester: Spring 2018 | −0.29 | 0.24 | 0.00 | −1.23 | 1.000 |
| Random effects | Variance | SD | | | |
| Instructor\|Semester (Intercept) | 0.00 | 0.05 | | | |
| Term (Intercept) | 0.11 | 0.33 | | | |
| Residual | 0.86 | 0.93 | | | |

**Table 5.** Multilevel Model with Length of Discussion Entries (Word Count)
Outcome Predicted by Instructor and Semester

| Fixed effects | b | SE | df | t-value | p-value |
|---|---|---|---|---|---|
| (Intercept) | 0.07 | 0.08 | 0.00 | 0.88 | 1.000 |
| Instructor: Instructor 2 | 0.27 | 0.06 | 603.00 | 4.52 | < 0.001 |
| Instructor: Instructor 3 | 0.18 | 0.05 | 603.00 | 3.28 | 0.001 |
| Instructor: Instructor 4 | 0.38 | 0.08 | 603.00 | 4.85 | < 0.001 |
| Instructor: Instructor 5 | 0.49 | 0.05 | 603.00 | 9.11 | < 0.001 |
| Semester: Spring 2018 | 0.06 | 0.08 | 0.00 | 0.80 | 1.000 |
| Random effects | Variance | SD | | | |
| Instructor\|Semester (Intercept) | < 0.01 | < 0.01 | | | |
| Term (Intercept) | 0.02 | 0.01 | | | |
| Residual | 0.48 | 0.70 | | | |

**Table 6.** Multilevel Model with Final Grade Outcome Predicted by Page Views, Discussion Posts (Count),
Semester, Discussion Posts × Instructor, and Page Views × Instructor

| Fixed effects | b | SE | df | t-value | p-value |
|---|---|---|---|---|---|
| (Intercept) | 91.19 | 3.29 | 3.77 | 27.68 | < 0.001 |
| Page views | 0.23 | 0.32 | 575.43 | 0.70 | 0.483 |
| Length of discussion entries | 5.85 | 0.46 | 551.05 | 12.62 | < 0.001 |
| Semester: Spring 2019 | −0.46 | 0.28 | 575.11 | −1.65 | 0.100 |
| Length of discussion entries × Instructor: Instructor 2 | −2.52 | 0.81 | 578.82 | −3.11 | 0.002 |
| Length of discussion entries × Instructor: Instructor 3 | −1.52 | 0.69 | 577.66 | −2.19 | 0.029 |
| Length of discussion entries × Instructor: Instructor 4 | 3.72 | 1.06 | 578.99 | 3.50 | 0.001 |
| Length of discussion entries × Instructor: Instructor 5 | −2.96 | 0.69 | 577.28 | −4.29 | < 0.001 |
| Page views × Instructor: Instructor 2 | 0.73 | 0.59 | 575.19 | 1.25 | 0.211 |
| Page views × Instructor: Instructor 3 | 0.20 | 0.46 | 575.17 | 0.44 | 0.662 |
| Page views × Instructor: Instructor 4 | −0.69 | 0.80 | 575.59 | −0.86 | 0.392 |
| Page views × Instructor: Instructor 5 | −0.39 | 0.57 | 574.98 | −0.69 | 0.492 |
| Random effects | Variance | SD | | | |
| Instructor (Intercept) | 53.10 | 7.29 | | | |
| Residual | 39.51 | 6.29 | | | |

**Table 7.** Multilevel Model with Final Grade Outcome Predicted by Page Views, Length of Discussion Entries (Word Count), Semester, Length of Discussion Entries × Instructor, and Page Views × Instructor

| Fixed effects | b | SE | df | t-value | p-value |
|---|---|---|---|---|---|
| (Intercept) | 65.60 | 5.26 | 3.76 | 12.47 | < 0.001 |
| Page views | 0.41 | 0.30 | 575.96 | 1.37 | 0.170 |
| Discussion posts count | 0.73 | 0.05 | 578.55 | 14.64 | < 0.001 |
| Semester: Spring 2019 | −1.25 | 0.28 | 577.58 | −4.43 | < 0.001 |
| Discussion posts count × Instructor: Instructor 2 | −0.21 | 0.08 | 200.07 | −2.47 | 0.015 |
| Discussion posts count × Instructor: Instructor 3 | −0.21 | 0.07 | 381.97 | −3.07 | 0.002 |
| Discussion posts count × Instructor: Instructor 4 | 0.36 | 0.09 | 155.32 | 3.87 | < 0.001 |
| Discussion posts count × Instructor: Instructor 5 | −0.20 | 0.08 | 203.64 | −2.41 | 0.017 |
| Page views × Instructor: Instructor 2 | 0.40 | 0.54 | 576.81 | 0.74 | 0.458 |
| Page views × Instructor: Instructor 3 | −0.22 | 0.44 | 576.03 | −0.51 | 0.609 |
| Page views × Instructor: Instructor 4 | 0.30 | 0.74 | 576.31 | 0.40 | 0.689 |
| Page views × Instructor: Instructor 5 | −0.45 | 0.54 | 575.96 | −0.84 | 0.404 |

| Random effects | Variance | SD |
|---|---|---|
| Instructor (Intercept) | 124.02 | 11.14 |
| Residual | 35.98 | 6.00 |

**Table 8.** Multilevel Model with Final Grade Outcome Predicted by Page Views during the First Two Weeks, Length of Discussion Entries (Word Count) during the first Two Weeks, Semester, Length of Discussion Entries × Instructor, and Page Views × Instructor

| Fixed effects | b | SE | df | t-value | p-value |
|---|---|---|---|---|---|
| (Intercept) | 90.65 | 1.66 | 3.77 | 54.52 | 0.000 |
| Page views | 0.51 | 0.39 | 589.14 | 1.33 | 0.184 |
| Length of discussion entries | 2.79 | 0.45 | 484.73 | 6.25 | 0.000 |
| Semester: Spring 2019 | −0.22 | 0.32 | 587.51 | −0.69 | 0.493 |
| Length of discussion entries × Instructor: Instructor 2 | −1.97 | 0.79 | 583.56 | −2.48 | 0.013 |
| Length of discussion entries × Instructor: Instructor 3 | −0.42 | 0.72 | 582.07 | −0.58 | 0.564 |
| Length of discussion entries × Instructor: Instructor 4 | 2.54 | 1.12 | 590.28 | 2.28 | 0.023 |
| Length of discussion entries × Instructor: Instructor 5 | −0.80 | 0.63 | 569.69 | −1.26 | 0.207 |
| Page views × Instructor: Instructor 2 | 0.92 | 0.69 | 588.13 | 1.32 | 0.186 |
| Page views × Instructor: Instructor 3 | 0.17 | 0.53 | 588.16 | 0.32 | 0.749 |
| Page views × Instructor: Instructor 4 | 0.17 | 1.05 | 589.46 | 0.16 | 0.874 |
| Page views × Instructor: Instructor 5 | −1.09 | 0.63 | 587.74 | −1.72 | 0.086 |

| Random effects | Variance | SD |
|---|---|---|
| Instructor (Intercept) | 12.90 | 3.59 |
| Residual | 53.35 | 7.30 |

**Sample Statistical Code**

The following sample code was run to produce the multilevel models in the previous section and in the main paper. The R version was 3.5.2, the Rstudio version was 1.2.1335, and it was run on a PC with Windows 10.

```
##### libraries #####
library(ggplot2) #for plotting interactions
library(ggsci) #for uchicago color palette
library(dplyr) #for data manipulation
library(tidyr) #for data manipulation
library(forcats) #for data manipulation
library(lme4) #for multilevel models
library(afex) #for take-one-out comparisons of fixed effects for multilevel models
library(emmeans) #for estimated marginal means of planned comparisons
library(multcomp) #for significance of estimated marginal means

##### multilevel models #####
#variable names:
#Final.Score = final course grade (%)
#pv = page views during the terms
#dl_all.wc = word count sum of all discussion entries
#disc_all_count = count of all discussion entries
#AddDrop = page views during the first two weeks
#dl_ad.wc = word count sum of discussion entries during first two weeks
#Instructor = categorical variable representing 5 instructors (reference = Instructor 1)
#Term = categorical variable representing term of instruction (reference = Fall 2018)
#df = dataframe containing all of these variables in raw form

#First, all predictor variables are transformed into a z-score (i.e., standardized and centered) due to large differences in scale
and because centering helps with interpretation of multilevel models.

df[, c("AddDrop","pv" ,"dl_all.wc","dl_ad.wc ", "disc_all_count")] <- scale(df[, c("AddDrop","pv" ,"dl_all.wc","dl_ad.wc ",
"disc_all_count")])

### Instructor and Semester as predictors ###

## outcome = final grade ##
#use the mixed() function in afex package to run a linear mixed effects model
fg_m <- mixed(Final.Score ~ Instructor + Term + (1|Term/Instructor), data = df)

#compare the main effects of each predictor by comparing model with and without that #predictor, via an F-test
nice(fg_m)

#full model output
summary(fg_m)

#looks like there is a marginal effect of Instructor. Let's do the pairwise comparisons using the #estimated marginal means (ak
least-squares means) package emmeans, using a tukey p-value #correction for multiple comparisons
emm_fg_m <- emmeans(fg_m, "Instructor")

#first get the estimated outcome (i.e., final grade) per instructor, then test differences among #instructors
emm_fg_m
update(pairs(emm_fg_m), by = NULL)

## outcome = page views ##
#use the mixed() function in afex package to run a linear mixed effects model
pv_m <- mixed(pv ~ Instructor + Term + (1|Term/Instructor), data = df)
```

#compare the main effects of each predictor by comparing model with and without that #predictor, via an F-test
nice(pv_m)

#full model output
summary(pv_m)

#looks like there is a significant effect of Instructor. Let's do the pairwise comparisons using the #estimated marginal means (ak least-squares means) package emmeans, using a tukey p-value #correction for multiple comparisons
emm_ pv_m <- emmeans(pv_m, "Instructor")

#first get the estimated outcome (i.e., final grade) per instructor, then test differences among #instructors
emm_ pv_m
update(pairs(emm_ pv_m), by = NULL)


## outcome = discussion entries length ##
#use the mixed() function in afex package to run a linear mixed effects model
dl_m <- mixed(dl_all.wc ~ Instructor + Term + (1|Term/Instructor), data = df)

#compare the main effects of each predictor by comparing model with and without that #predictor, via an F-test
nice(dl_m)

#full model output
summary(dl_m)

#looks like there is a significant effect of Instructor. Let's do the pairwise comparisons using the #estimated marginal means (ak least-squares means) package emmeans, using a tukey p-value #correction for multiple comparisons
emm_ dl_m <- emmeans(dl_m, "Instructor")

#first get the estimated outcome (i.e., final grade) per instructor, then test differences among #instructors
emm _dl_m
update(pairs(emm_ dl_m), by = NULL)


### Add in LMS behavior predictors ###

### discussion post number ###
#use the mixed() function in afex package to run a linear mixed effects model
dc_m <- mixed(Final.Score ~ pv+disc_all_count + disc_all_count:Instructor + pv:Instructor + Term + (1|Instructor), data = df)

#compare the main effects of each predictor by comparing model with and without that #predictor, via an F-test
nice(dc_m)

#full model output
summary(dc_m)

#there was a significant effect of Term, so let's first estimate the model-predicted final grade for each term, across instructors and for the average page views and number of discussion posts.
emm_ dc_m <- emmeans(dc_m, "Term")
emm_ dc_m
#now test those differences in model-estimated final grade in the different terms
update(pairs(emm_ dc_m))

#Since there was a significant interaction of discussion post number and instructor, let's get the #estimates and comparisons of slopes of the effect of discussion post number on final grade for #each instructor, across term and for average page views
emtrends(dc_m, pairwise ~ Instructor, var = "disc_all_count")

### discussion entries length (word count) ###
#use the mixed() function in afex package to run a linear mixed effects model
dl_all_m <- mixed(Final.Score ~ pv+dl_all.wc + dl_all.wc:Instructor + pv:Instructor + Term + (1|Instructor), data = df)

#compare the main effects of each predictor by comparing model with and without that #predictor, via an F-test
nice(dl_all_m)

#full model output
summary(dl_all_m)

#Since there was a significant interaction of discussion entries length and instructor, let's get the #estimates and comparisons of slopes of the effect of discussion post number on final grade for #each instructor, across term and for average page views
emtrends(dl_all_m, pairwise ~ Instructor, var = "disc_all_count")

##### first two weeks #####

#use the mixed() function in afex package to run a linear mixed effects model
tw_m <- mixed(Final.Score ~ AddDrop+dl_ad.wc + dl_ad.wc:Instructor + AddDrop:Instructor + Term + (1|Instructor), data = df)

#compare the main effects of each predictor by comparing model with and without that #predictor, via an F-test
nice(tw_m)

#full model output
summary(tw_m)

#Since there was a significant interaction of discussion entries length and instructor, let's get the #estimates and comparisons of slopes of the effect of discussion post number on final grade for #each instructor, across term and for average page views
emtrends(tw_m, pairwise ~ Instructor, var = "dl_ad.wc")


**Heat Maps**

Section 3.1 of the main paper presents heat maps with 10 randomly sampled students per instructor and per semester. The random sampling had no methodological purpose other than making the graphs easier to interpret, with the main takeaway being that there were more student page views in the first half of the semester. That is, the page views of 50 students are currently shown instead of all 332 for the fall semester and 227 for the spring semester. We present here the heat maps of all students per semester.
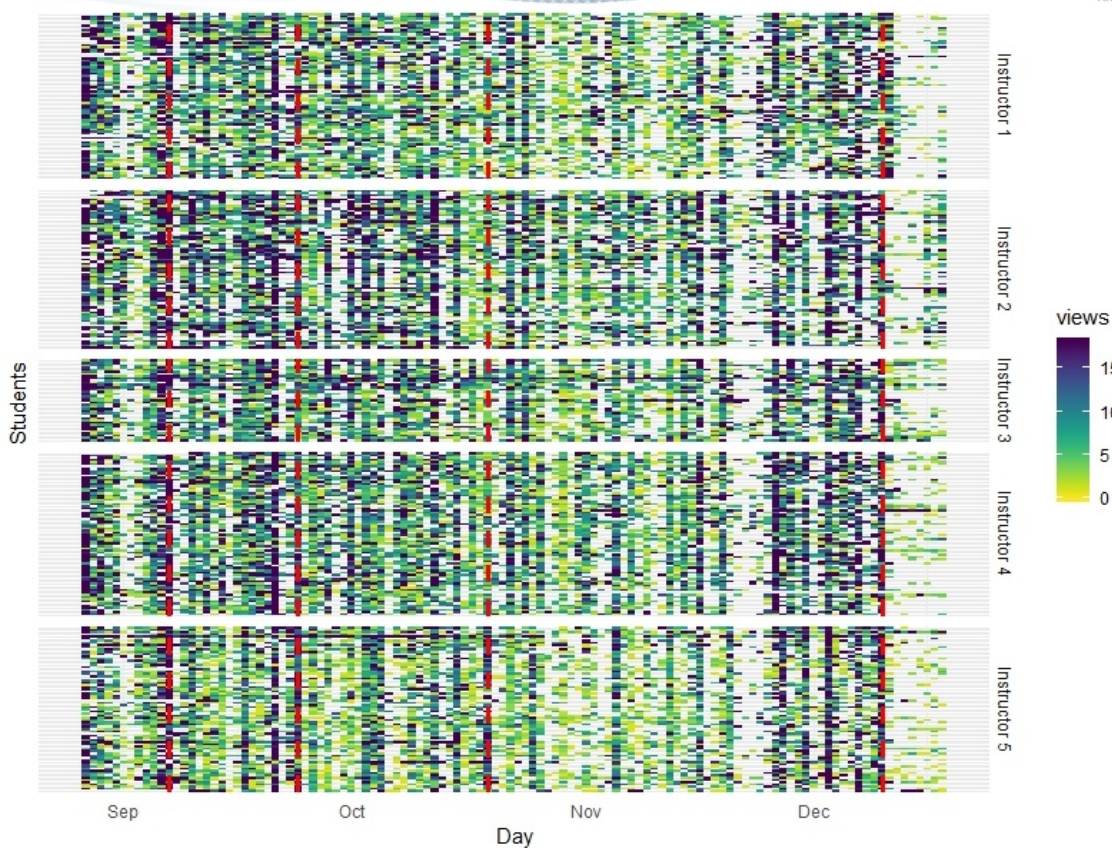
**Figure 1.** Example fall 2018 course activity for all students per instructor, corresponding to Figure 2 in the main text
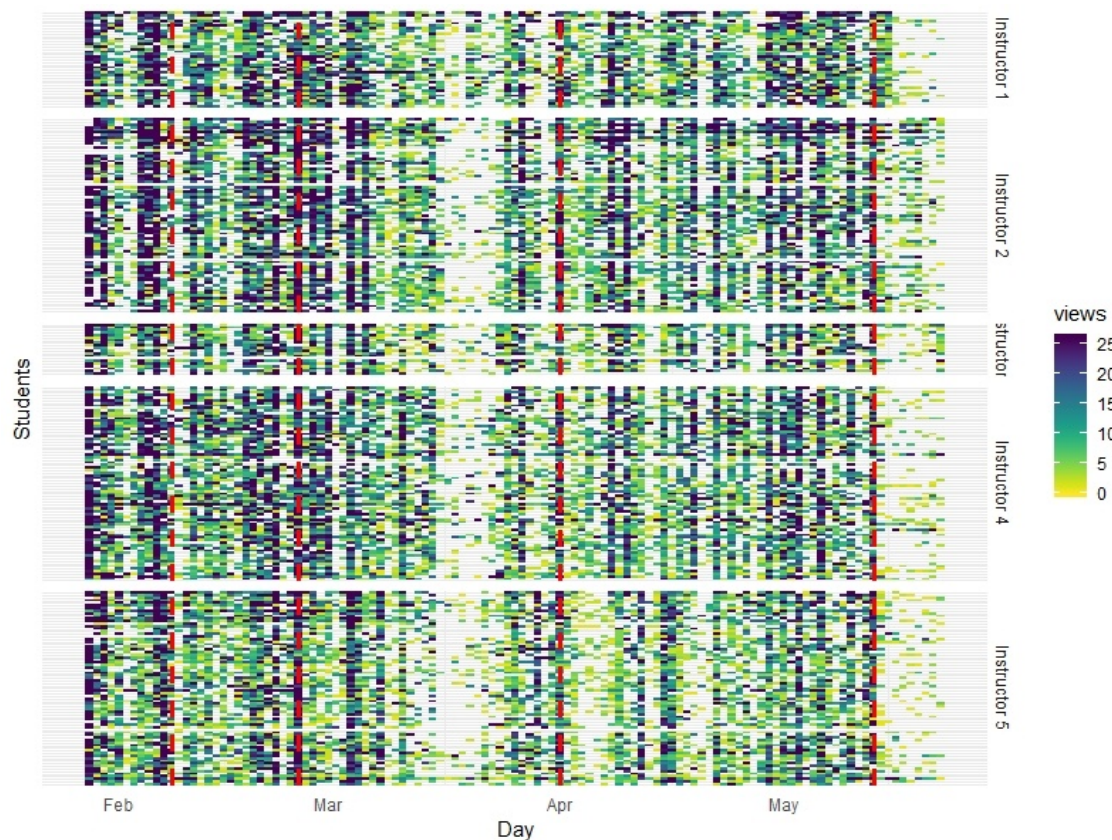


**Figure 2.** Example spring 2019 course activity for all students per instructor, corresponding to Figure 3 in the main text