What can Digital Trace Data Tell Us about Post-secondary Students' Academic Success?

An Overview of the Literature and an Illustrative Example

Luise von Keyserlingk, Fani Lauermann, Renzhe Yu, Charlott Rubach, Richard Arum

Abstract

Self-regulated learning (SRL) is an important facilitator of students' academic success in post-secondary education. We provide an overview of the growing literature that uses digital trace data to investigate students' study behaviors and SRL in post-secondary institutions. Digital trace data such as (changes in) click activity obtained from learning management systems (LMS; e. g., Moodle, Ilias, or Canvas) can be a useful indicator of SRL regarding students' time management and aspects of monitoring behavior. Whereas broad measures of study activities in online environments, such as the number of clicks and time spent on course activities do not consistently predict performance, more fine-grained measures, such as number of clicks before deadlines in courses or using voluntary quizzes to monitor learning, can identify adaptive self-regulated learning strategies. Hereby, information about course design and context is essential for creating such SRL measures. In addition, multi-source data (e. g., digital traces and self-reported learning strategies) are needed to capture not only behavioral but also motivational and meta-cognitive aspects of SRL. In this study, we provide an illustrative example of the type of digital trace data that can be collected via LMS to predict students' academic success using data from the ongoing longitudinal UCI-MUST project.

Keywords: self-regulated learning; digital trace data; post-secondary education; empirical longitudinal data; course performance

Zusammenfassung

Selbstreguliertes Lernen (SRL) ist ein wesentlicher Prädiktor für akademischen Erfolg in postsekundärer Bildung. Das Kapitel gibt einen Überblick über aktuelle Studien, die digitale Verhaltensspurdaten zur Untersuchung von Lernverhalten und SRL von Studierenden nutzen. Digitalen Verhaltensspuren, die zumeist von Lernmanagement Systemen wie Moodle, Ilias oder Canvas gewonnen werden, eignen sich insbesondere für die Untersuchung des Zeitmanagements oder der Selbstevaluation von Studierenden als zentrale Aspekte von SRL. Dabei zeigt sich, dass quantitative Maße von Lernverhalten, wie die gesamte Anzahl an Klickaktivitäten oder die Dauer die Studierende auf einem online Kurs verbringen eher moderate Prädiktoren von Kursleistung darstellen. Spezifischere Maße von SRL, wie z. B. Veränderungen im Klickverhalten von Studierenden vor Deadlines oder die Nutzung regelmäßiger freiwilliger Testfragen zur Überprüfung des Lernzuwachses sind hingegen vielversprechendere Maße um adaptive und erfolgreiche SRL Strategien zu erfassen. Hierbei sind detaillierte Informationen über den Kursablauf und -kontext zu berücksichtigen. Um neben diesen vornehmlich deskriptiven Maßen von SRL auch internale SRL Prozesse identifizieren zu können, sollten zusätzlich zu digitalen Verhaltensspuren auch Selbstberichtdaten zu motivationalen und meta-kognitiven SRL Elementen erhoben werden. Neben einem Überblick über aktuelle Studien nutzen wir Daten der UCI-MUST Studie, um ein Beispiel zur Verwendung von digitalen Verhaltensspuren zur Beschreibung von Lernverhalten und zur Vorhersage von Kursleistungen von Studierenden zu geben.

Schlagworte: Selbstreguliertes Lernen; digitale Verhaltensspurdaten; postsekundäre Bildung; empirische längsschnittliche Daten; Kursleistung

1. Introduction

The ability to plan, monitor, and adjust one's own study activities is important for students' learning success. Particularly in higher education, where learning environments are less structured than in secondary school, and students are afforded more autonomy to choose their courses and structure academic activities, self-regulated learning (SRL) skills, such as planning, monitoring, and adjustment of study behaviors, are particularly important predictors of students' academic success (Broadbent & Poon, 2015; Zimmerman & Schunk, 2011). The challenging years since the outbreak of the COVID-19 pandemic have further shown that SRL-skills are central facilitators of college students' academic success (Berger et al., 2021; Juriševič et al., 2021). The closure of university campuses and the shift to remote learning during the pandemic drastically changed the learning environments of college students. While structures such as synchronous face-to-face classes, libraries, or study groups on campus were no longer available, students had to develop new study routines with remote learning activities. Learning management systems (LMS), such as Moodle, ILIAS, and Canvas, are designed to facilitate teaching and learning and have become a central element of instruction in post-secondary education settings. During the pandemic, LMS provided a key platform for remote teaching and learning (e. g., for sharing study materials, assignments, and videotaped classes). Moreover, LMS also provide new and promising approaches to investigate students' SLR and learning outcomes in authentic educational contexts. For example, through observations of students' authentic interactions with study materials provided via LMS and their study behaviors such as time management, adherence to deadlines, and click activities.

Self-regulation research has a long tradition of using self-report data to investigate students' SRL (e. g., Pintrich et al., 1991; Weinstein & Palmer, 2002). An advantage of such self-reported data is that they can capture internal cognitive, motivational, and affective processes that are not directly observable in students' behavior (e. g., aspirations and goals, elaboration strategies, outcome expectations, or anxiety and enjoyment during study activities). Furthermore, surveys with established SRL instruments can be administered at relatively low costs and can be applied to any educational context (Wolters & Won, 2018). However, students' self-reports of their study behaviors and learning strategies can be subject to various reporting biases, for instance, due to insufficient memory or the elicitation of socially desirable responses (Baker et al., 2020). Accordingly, self-regulation researchers often rely on multiple data sources to obtain a more comprehensive and precise assessments of SRL. Since the implementation of LMS in educational contexts, SRL researchers increasingly use digital trace data from these LMS to examine students' learning behavior in different courses (Baker et al., 2020; Crompton et al., 2020; Li et al., 2020). Different measures, such as the number of clicks, number of study sessions, or time spent in a course per day can be used to quantitatively describe students' study behaviors, engagement, and learning patterns in a given course. In this chapter, we provide an overview of the growing literature on digital trace data that can be obtained via LMS to investigate students' learning behavior and success. In particular, we focus on studies using digital trace data to measure students' SRL behaviors in post-secondary education, and we outline some of the key advantages and challenges of using such data in SRL research. Second, we use data from the ongoing longitudinal UCI-MUST project (Arum et al., 2021) to provide an illustrative example of the type of digital trace data that can be collected via LMS to predict students' academic success and potentially aid their SRL.

2. Self-regulated Learning in College

In post-secondary education, students are often required to navigate a challenging curriculum and organize their study activities in several courses that they take simultaneously. Thus, SRL is a central skill that facilitates learning and success in post-secondary education (Broadbent & Poon, 2015). SRL entails multiple cognitive, motivational, behavioral, and affective components, and existing theoretical models on SRL describe and integrate these components in somewhat different ways. Boekaerts (1999), for instance, describes a set of skills and resources that are central to self-regulated learning in a multi-layered model, whereas Zimmerman (1990) describes SRL as a cyclical process where learners use different sets of SRL skills and behaviors in a structured and recurrent manner. Boekaerts' multi-layered model suggests that SRL behaviors unfold on different levels of proximity to the learning content. These include the choice of adaptive cognitive strategies to process the *learning material*, the use of meta-cognitive strategies to regulate the *learning process*, and the choice of suitable goals and recourses to facilitate learning in specific *environments* (Boekaerts, 1999, 2010). Process models, by comparison, describe three central elements of self-regulated learning that occur cyclically: (a) *forethought*, (b) *performance and volitional control*, and (c) *self-reflection* (Zimmerman, 1990). The phase of *forethought* or *planning* entails the setting of specific learning goals and the selection of learning strategies. The second phase of *performance and volitional control* requires students to apply the selected learning strategies and monitor and control their learning process. The phase of *self-reflection* includes self-evaluation and causal attribution of the learning outcomes, for instance, to factors that are internal and controllable or external and not controllable by the student, and informs subsequent planning phases (Zimmerman, 1990; Zimmerman & Schunk, 2011).

Although both theories emphasize different aspects of SRL, they have a large overlap in terms of the described cognitive, meta-cognitive, motivational, and behavioral skills that are essential to SRL. Depending on the chosen theoretical framework, empirical SRL research often focuses either on cyclical aspects of SRL behaviors or on specific cognitive and meta-cognitive skills and strategies of learners. Extensive literature shows that students with higher SRL skills attain better learning outcomes and are more successful in college (Broadbent & Poon, 2015; Dörrenbächer & Perels, 2016; Kitsantas et al., 2008; Zimmerman, 1990). Most of this literature used well-established survey instruments to measure students' SRL (e. g., Motivated Strategies for Learning Questionnaire – MSLQ: Pintrich et al., 1991; Learning and Study Strategies Inventory – LASSI: Weinstein & Palmer, 2002). Over the past years, however, there has been growing interest in the use of digital trace data from LMS, besides established survey instruments, as a means to investigate SRL (Arizmendi et al., 2022; Bernacki et al., 2020; Cogliano et al., 2022). Such behavioral trace data provide complementary data on students' study behaviors, in addition to their self-reports. However, an open question that warrants careful consideration concerns the interpretability of digital trace data as an indicator of central SRL components, such as cognitive and meta-cognitive skills described by Boekaerts (1999; 2000), or cyclical aspects of SRL behavior described by Zimmerman (1990; 2011).

2.1. Measuring Self-regulated Learning with Digital Trace Data

Digital trace data from learning management systems capture students click activities in log files and, thus, such data allow capturing students' study behavior in authentic contexts (Arizmendi et al., 2022; Bernacki et al., 2020; Crompton et al., 2020). Digital traces of students' study behaviors provide an opportunity to measure certain aspects of self-regulated learning, such as time management, regularity of study efforts, and self-testing (e. g., using voluntary quizzes or assignments to test one's knowledge of course contents). Other central elements of self-regulated learning that do not manifest directly in observable behavior, such as goal-setting, cognitive strategy use (e. g., elaboration of learning material), and emotion regulation are comparatively more difficult to measure with such data (Bernacki, 2018). Current research can be categorized broadly into a) studies that focus on global measures of study behavior, such as the overall number of clicks and time spent on courses, and b) studies that investigate specific aspects of self-regulated behavior, such as active planning behavior, cramming versus spacing, self-testing behavior, help-seeking behavior, and others. The main interest of this chapter is to provide an overview of different approaches that have been used to measure students' learning behaviors and different aspects of SRL with digital trace data. The main objective of this chapter is not to provide an exhaustive review of the existing literature, but rather to review recently developed and applied approaches to measuring central aspects of SRL behaviors with digital trace data in post-secondary education settings.

3. Links Between Digital Trace Data and Self-regulated Learning Behaviors in Post-secondary Education

3.1. Global Measures of Study Behavior Based on Digital Traces

Several relatively global measures of study behavior can be obtained from digital traces in learning management systems (LMS). Commonly used global assessments include students' overall number of clicks when using the course site via the LMS, time spent navigating through and interacting with course material online, and the number of online study sessions in a course (Baker et al., 2020; Cicchinelli et al., 2018; Greene et al., 2021). Such measures describe the quantity of students' study actions but do not capture the quality or type of study actions. Nevertheless, such measures can predict desirable learning outcomes. For instance, Cicchinelli et al. (2018) found that first-year students in a computer science program who produced more clicks overall had more study sessions, had longer durations of interacting with the course site in the online LMS that was used for their lecture, and attained better grades in quizzes and final exams in the lecture. Similarly, in a distance-learning university in the UK, Nguyen et al. (2020) investigated the associations between the time students spent on their courses online and their course performance. Using data from roughly 150,000 students, they showed that students who spent more time studying for their courses – as inferred by how much time they spent interacting with course contents on LMS – were more likely to pass their exams and obtained better grades. While these associations were observed for all students in the courses, Nguyen et al. (2020) described differences in study activities for students with different demographic backgrounds. On average, female students spent more time on their college courses on LMS compared to their male peers, whereas students from underrepresented ethnicities spent less time on LMS course content and obtained lower course grades. Authors provide different explanations for these results: Students of underrepresented minorities might have competing obligations, such as work in addition to studying, that could constrain their available time and resources for studying. In addition, these students might lack knowledge about effective study and SRL strategies and may therefore show less adaptive study behaviors.

Broad measures, such as the number of clicks, study sessions, and time spent on a course provide additional insights into students' self-regulated learning when observed on a daily or weekly aggregation level across a course. With such measures, researchers can investigate changes in the number of study activities across specific time periods, the regularity, and the fluctuation in students' study activities. Park et al. (2017), for instance, identified three patterns of changes in students' click activities across the duration of an online and a face-to-face course with LMS at a public university in the US. They categorized students' study behaviors into "increasing click activity," "no change," or "decreasing click activity" across the courses. Students who had an increasing pattern of click activity across the course were more likely to pass the course than students who had a decreasing pattern. Focusing on students' weekly click activity in a course where students regularly received tasks on Mondays with a submission deadline on Fridays, Park et al. (2018) showed that students with a more regular click pattern working on the tasks throughout the week obtained better course grades compared to students who had increased click activity only shortly before the deadline on Fridays. These studies suggest that students who maintain more regular and continuous study activities tend to attain better performance outcomes.

While the above-mentioned studies reported positive associations between broad measures of study activities on LMS and course performance, other studies suggest no or only very small positive associations between the overall number of click activity and time spent on the course sites with course performance outcomes (see, e. g., Greene et al., 2021, You et al., 2016). Mixed findings could be explained by different course designs and types of use of LMS by the instructors. Furthermore, mixed findings could be related to different levels of granularity of behavioral trace measures ranging from variables on the course level (e. g., time spent on the course across the entire semester (Nguyen et al., 2020), daily click activities across the entire semester (Park et al., 2017), or daily click activities per week (Park et al., 2018)). Decisions about the level of aggregation of behavioral trace data should be driven by the research question and available information about the course context. For instance, only if information about the course design and course deadlines is available, researchers can meaningfully interpret increases in study activities before certain dates or can derive measures such as 'time to a deadline' to investigate aspects of students' time management in their courses.

3.2. SRL-specific Measures of Study Behavior Based on Digital Traces

Digital trace data can also be used to infer specific types of self-regulated behaviors, as conceptualized by Boekaerts (1999, 2010) and Zimmerman (1990, 2000). These include, for instance, measures of time management and regularity of study activities that can be linked to cyclical processes of SRL (Zimmerman, 1990; 2000), and to the successful use of meta-cognitive strategies to regulate learning processes (Boekaerts, 1999; 2010). Furthermore, such measures can describe the use of specific meta-cognitive strategies, such as help-seeking behaviors, monitoring of learning outcomes with self-tests and quizzes or monitoring performance outcomes through accessing grade books in a course. We provide an overview of studies that used such measures in the following sections.

Time Management and Procrastination Measured with Digital Trace Data

When relevant information about the course context and course design is available, researchers can generate more SRL-specific variables with students' digital trace data to investigate SRL and performance in college courses. Information about deadlines and due dates, for example, can be used to generate measures for procrastinating behaviors (Li et al., 2020; Rodriguez et al., 2021). Such measures can include the time between the submission of an assignment and the submission deadline and the proportion of assignments and content material accessed before versus on a due date. Using such measures of self-regulated behaviors from an online college course, Li et al. (2020) showed that students who proportionally accessed more study units before the due date than on the due date, and who submitted assignments longer in advance of the deadline, were students who obtained better final course grades. Similarly, Rodriguez et al. (2021) investigated whether regular access versus irregular and delayed access of lecture videos predicted final course grades in an asynchronous online course. The course contained 48 short lecture videos divided into four modules. In each module, students needed to watch the corresponding lecture videos in a pre-specified order until a specific due date. Rodriguez et al. (2021) identified four clusters of study behaviors with students who a) watched nearly all videos before the due date (early planners), b) watched most videos before the due date and only a few on the due date (planners), c) watched most videos on the due date (procrastinators), and d) watched only a few videos and all of them late (low engagers). Students who were 'early planners' and 'planners' obtained better course grades than their peers who were identified as 'procrastinators' or 'low engagers'. Rodriguez et al. (2021) further investigated if students with certain background characteristics were at particular risk for belonging to a cluster with maladaptive study patterns (i. e., procrastinators or low engagers). Results showed that students of low-income families and first-generation college students more often had low engagement patterns in their courses, compared to their peers, and obtained lower grades in their courses.

Monitoring and Self-evaluation Measured with Digital Trace Data

Digital trace data can also be used to identify study behaviors that are related to planning, monitoring, and self-evaluation, depending on what supplemental information is available about the course (see, e. g., Greene et al., 2021; Huang et al., 2022). For instance, students' use of course calendar functions and visits of course modules that show the course syllabus at the beginning of the course and before course exams can provide information about *planning* behaviors. Course syllabi are required for each college course in the US and typically provide information about course activities, requirements, and grading policies in the course. Students' completion of voluntary quizzes in the course (i. e., opportunities to self-test course content and evaluate knowledge gaps or learning gains) can indicate *monitoring* of the learning progress. Students' regular access of (online) performance feedback from the instructors and gradebooks can indicate *self-evaluation* practices. Latent profile analyses have been used to investigate to what extent students show *planning*, *monitoring*, and *self-evaluating* behaviors when they use course materials provided via LMS (Greene et al., 2021; Hong et al., 2020; Huang et al., 2022; Li & Baker, 2018). Students who regularly participated in ungraded and/or voluntary quizzes attained higher course grades than their peers who showed less quiz-taking behavior (Carvalho et al., 2022; Ifenthaler et al., 2022; Huang et al., 2022; Li & Baker, 2018). A possible explanation provided by the authors is that students who regularly participated in the quizzes were thus able to monitor their learning progress and adjust learning strategies as needed. Greene et al. (2021), for instance, used data from 408 students enrolled in a biology class. They used latent profile analysis to identify SRL behaviors related to planning activities (e. g., accessing the course syllabus and using the course calendar, as well as reading announcements), information acquisition (e. g., attending class meetings, accessing additional course readings), and help-seeking (e. g., reaching out for help, clicking on links to learning support services). Students who showed more planning activities and information acquisition obtained better course grades than their peers. Similarly, Hong et al. (2020) used latent profile analysis to investigate if students predominantly used SRL strategies related to planning (by visiting the course syllabus and study guides), monitoring of their learning progress (by taking regular exercises and quizzes), and monitoring and evaluating their performance (by visiting their gradebooks). The sample consisted of digital trace data from 1,326 college students in biology classes at a mid-western university in the US. Most students showed little planning and monitoring behaviors. About 15 % of the students showed more planning behaviors and frequently monitored their performance by visiting the course gradebook. About 10 % of the enrolled students frequently monitored their learning through quizzes. Students who regularly showed behaviors related to monitoring their learning and performance through quizzes and gradebooks outperformed their peers and obtained higher final course grades. These findings indicate that students who showed more study activities that can be linked to critical aspects of SRL were more successful in their courses.

However, the above-described findings derived from digital trace data remain on a rather descriptive level of observable study behavior and associations with desirable performance outcomes. An open question is, whether students enact certain study behaviors, such as regular completion of self-tests and quizzes, because they are encouraged or required to do so by their instructor (i. e., *externally regulated behavior*), or because they voluntarily and purposely used this strategy to monitor their learning progress (i. e., *self-regulated behavior*). Furthermore, these studies did not examine whether students' SRL activities related to monitoring and self-evaluation were predictive of subsequent changes in study behaviors and course performance. A promising approach to further distinguish between self-regulated and externally regulated study activities in (online) course environments is to combine behavioral trace data with (a) self-reported SRL behaviors, and (b) pertinent information about students' learning context (e. g., course requirements). The combination of different sources of information about students' study activities would enable analyses of whether students' study behaviors, as observed via digital trace data, are driven by students' SRL skills and purposefully selected learning strategies, by course requirements, or by a combination of both.

Hence, a relevant question for SRL researchers is to what extent self-reported data and digital trace data on self-regulation overlap, diverge, or complement each other in predicting performance and improving our understanding of SRL in authentic contexts (Baker et al., 2020; Bernacki, 2018). Some of the above-mentioned studies used digital trace data and survey data to investigate students' SRL behaviors in post-secondary education. Results on correlations between self-reported SRL skills and SRL behaviors measured with digital trace data are mixed. Huang et al. (2022) showed that both self-reported self-efficacy and the use of metacognitive strategies (i. e., planning and monitoring) measured with digital trace data predicted course grades. However, the two types of measures were not significantly correlated. Similarly, Cicchinelli et al. (2018) found no significant correlations between students' self-efficacy and overall study activity, time spent on coursework, and monitoring and planning activities. However, they found moderate positive correlations between students' self-reported self-regulation skills and self-regulation measures derived from digital trace data. Li et al. (2020) assessed self-reported self-regulation skills with surveys at the beginning (T1) and end (T2) of a quarter. Time management measured with digital trace data in online lectures correlated positively with self-reported self-regulation skills measured only at T2. Ifenthaler et al. (2022) focused on self-testing strategies and found that students who reported using more self-testing strategies in their courses also engaged in more self-assessment tasks in the LMS of their course. Although not entirely consistent, these findings point to positive associations between self-reported data and digital trace data on SRL skills. A key factor that may contribute to these inconsistencies is the timing of measurement and the level of generality of different types of measures. For instance, students' self-reported broader motivational beliefs such as generalized academic self-efficacy (Cicchinelli et al., 2018; Huang et al., 2022) and self-reported SRL assessed before their course had started (Li et al., 2020) are often not significantly related to students' digital trace data collected during the semester. In contrast, when students report on their SRL after they have already participated in the course for a few weeks, the associations between students' self-reported SRL and their digital trace data tend to be stronger (Cicchinelli et al., 2018; Li et al., 2020). These findings indicate that the time point and specific aspects of SRL (e.g., self-testing to monitor learning progress) measured with self-reported data and digital trace data should be aligned when these measures are being used complementarily.

4. Self-regulated Learning in College: An Illustrative Example of Using Digital Trace Data

In the second part of this chapter, we provide an example of how digital trace data can be used to describe students' study behaviors over time. Based on a collaboration between TU Dortmund and the University of California, Irvine, we were able to use data from the UCI-MUST project (Arum et al., 2021), an ongoing longitudinal study to examine undergraduates' experiences and factors that facilitate college success. We used data from undergraduate students who were enrolled in two large biology lectures in the fall of 2020 and were using course materials that were provided to them via LMS. Importantly, all students were studying remotely due to the COVID-19 pandemic and the social distancing rules that were in place at that time. Consequently, the LMS used at the time includes rich data on students' digital traces and learning behaviors. We focused on a selected subsample of students and examined the pattern of students' study activities across the ten weeks of the academic fall quarter. Our analyses focus on three key research questions:

1. Is variability in study activities on LMS across the quarter associated with critical course events (i. e., midterm exams)?
2. Are global and week-specific measures of study activities on LMS across the quarter associated with students' demographic background variables and final course grades?
3. Is students' self-reported self-efficacy for self-regulated learning related to their study activities on LMS across the quarter?

4.1. Sample and Procedure

We used data from *N1* = 805 undergraduate students who were enrolled in two large biology lectures in their junior year in the fall term of 2020. These biology lectures are usually face-to-face lectures and instructors use the LMS Canvas to provide course materials and assignments. In the fall term of 2020, the lectures were shifted to a fully remote format because of the COVID-19 pandemic. Canvas was used to provide lecture content and administer assignments and midterm exams. The sample consisted of a diverse student population with 44 % first-generation college students, 25 % students who belonged to a historically underrepresented minority (Latino, African American, Pacific Islander), and 67 % female students. We used different data sources to examine the proposed research questions.

Digital Trace Data. We used the overall number of clicks each student used per day in their Canvas course as a quantitative measure of study activity. This measure includes, for example, clicks on course materials, downloads of course material, uploads of assignments, and completion of quizzes and midterm exams. First, we aggregated daily click activities on a weekly level per student. Second, we centered students' weekly click activities at the course mean to remove between-course variance in the two biology lectures (within-course centering), and thus created comparable study activity measures in both lectures. We decided to aggregate daily click activities on a weekly level for two reasons: First, we were interested in the variability in students' click activity across the entire term and during significant week-specific course events (midterm and final exams). Second, the two lectures had the same structure and exams happened in the same weeks, but lectures and exams took place on different weekdays. Aggregation of click activities on a weekly level thus improved comparability of the two courses.

Course Syllabi. Course syllabi are detailed course plans that were available for both lectures. We used the course syllabi to identify the dates of midterm exams. In both lectures, midterm exams took place in weeks 3, 6, and 9 of the fall term of 2020.

Administrative Data. Data on students' demographic backgrounds and final course grades were obtained from students' college records. We used dichotomous variables as indicators of students' first-generation college-going student status (1 = yes; 0 = no), if students belonged to a historically underrepresented minority (URM) on campus (1 = yes; 0 = no), and about students' biological gender (1 = female; 0 = male). Administrative data included an option 'other' to declare students' gender, but all students in the present sample had a record of either female or male gender. Furthermore, we used high school grade point average (GPA) as an indicator of prior achievement. We used final grades as an indicator of course performance. Students received letter grades (A – F) in both lectures. We transformed the letter grade to a numeric variable (A = 12 to F = 0), with higher values indicating better performance.

Survey Data. A small subsample of the 805 students in the selected biology lectures participated in surveys during the UCI-MUST project. In the UCI-MUST project, more than 1.200 undergraduates from all fields of study consented to participate in the survey study of the project. Of those, 25 students were enrolled in the two biology lectures that are presented in the illustrative example of this chapter. These students completed a survey at the beginning of the fall 2020 quarter that included questions about their self-efficacy for self-regulated learning (T1; *N2T1* = 25), and 18 of these students completed the same questions again after the fall term of 2020 (T2; *N2T2* = 18). We used five items to measure self-efficacy for self-regulated learning that were based on the self-efficacy scale by Farr et al., (2011). Two example items are: “How good are you at motivating yourself to do schoolwork” and “How good are you at finishing your homework assignments by deadlines”. Students responded to the items on a slider scale from 0 – *not at all good* to 100 – *exceptional*. Because of the very small sample sizes, we treat findings of analysis with survey data as preliminary suggestive evidence that needs to be extended and continued with larger samples.

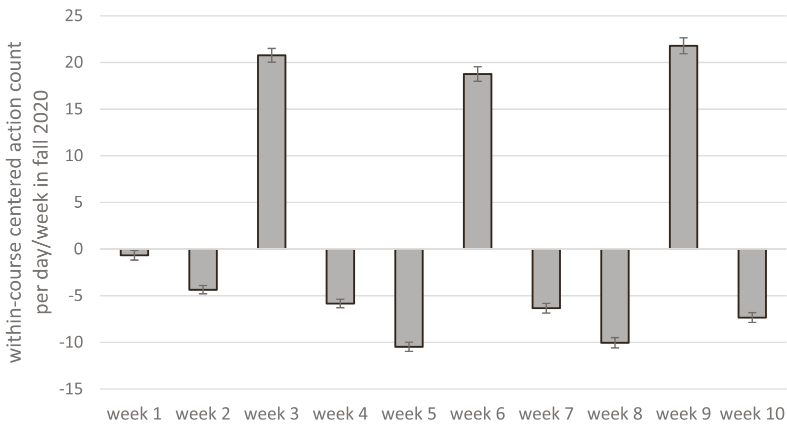
We used descriptive statistics and correlative analyses to describe the pattern of study activities across the quarter and the associations of study activities with demographic variables, course grades, and self-reported self-efficacy for self-regulated learning.

4.2. Results

Associations of Variability in Click Activities Across the Quarter with Critical Course Events (RQ1)

On average, students had 27 action counts on each day in a week in their course. A large standard deviation and a large range from minimum to maximum action counts per day indicate large variability in click activity across days of the quarter (min = 1, max = 734, M = 26.92, SD = 28.29). Figure 1 shows the pattern of click activity on a weekly level across the fall 2020 quarter. Overall, the pattern shows a small decline in click activities across the 10 weeks of the quarter with large increases in click activities in weeks 3, 6, and 9 compared to the mean click activity in the course, and decreased click activities in weeks 4, 5, 7, 8, and 10. Information from available course syllabi explained these fluctuations in click activities across the week: Click activity increased in weeks of midterm exams (weeks 3, 6, and 9), and decreased in weeks after the midterm exams. This finding emphasizes the importance of using relevant context information about the courses. Information on relevant deadlines and exams is central for a meaningful interpretation of study patterns.

Figure 1: Behavioral trace data from two biology lectures in fall 2020. N1 = 805 undergraduate students. Within-course centered action counts per day aggregated on a weekly level. Error bars indicate standard errors.



Associations of Click Activities with Students' Demographic Background and Course Performance (RQ2)

In a second step, we explored associations of click activities across the quarter with students' demographic backgrounds and their course performance. Table 1 shows the results of bivariate correlations of these variables. Overall, click activities differed only slightly by students' demographic backgrounds. Female students had slightly more click activities across the quarter (small positive correlations of female gender with fall 2020 overall study activity r = .09; and with study activity in weeks 1, 4, 5, 6, and 10 r = .07 to r = .11). Students of an underrepresented minority had slightly lower study activity in weeks of the midterm exams (r = -.10 to r = -.13). First-generation and continuing-generation college students did not differ systematically in their click activities across the quarter. Furthermore, high school GPA was not significantly associated with click activities during the quarter.

Students' final course grade had a small positive correlation with students' overall click activity across the quarter (r = .15) and small positive correlations in several weeks of the quarter (weeks 2 to 4: r = .11 to r = .17; weeks 6 to 9: r = .12 to r = .14). Hence, click activity in the weeks of the midterm exams, as well as the weeks before and after the midterm exams was significantly correlated with students' final course grades.

Table 1: Bivariate correlations of within-course centered action counts per week with demographic variables and final grade.

	f20 action count	week 1 action counts	week 2 action counts	week 3 action counts	week 4 action counts	week 5 action counts	week 6 action counts	week 7 action counts	week 8 action counts	week 9 action counts	week 10 action counts
Female	**0,09**	**0,08**	0,04	0,04	**0,09**	**0,07**	**0,07**	**0,11**	0,06	0,06	**0,07**
Underrepresented minoriy	-0,04	-0,01	-0,03	**-0,13**	0,00	0,05	**-0,10**	0,01	0,04	**-0,11**	-0,04
First-generation college student	0,01	0,04	0,04	-0,04	-0,03	0,03	-0,01	0,02	0,06	-0,05	0,02
High school GPA	0,03	-0,03	-0,01	0,04	0,01	-0,01	0,02	0,03	0,06	0,04	0,02
Final grade	**0,15**	0,04	**0,11**	**0,17**	**0,12**	0,06	**0,12**	**0,13**	**0,11**	**0,14**	0,05

Note. *N1* = 805 students. Within-course centered weekly action counts.
Greyed cells – weeks with midterm exam (week 3, 6, 9). Bolt font $p < .05$.

Associations Between Self-reported Data and Digital Trace Data (RQ3)

Finally, using data from a small subsample of students who participated in the UCI-MUST project surveys, we examined the associations between students' self-reported self-efficacy for SRL and students' click activities across the quarter (*N2T1* = 25; *N2T2* = 18). Results shown in Table 2 indicated that students with higher self-efficacy for self-regulation at the beginning of the fall 2020 quarter (T1) had higher click activities in several weeks across the quarter. This association was large and statistically significant at the beginning of the quarter and in the weeks around the first two midterm exams (weeks 1 to 3: $r = .41$ to $r = .50$; weeks 5 to 6: $r = .40$ to $r = .65$). The associations between students' self-efficacy for SRL measured at T2 (shortly after the fall 2020 term) and their study activities were positive, but not statistically significant in most weeks of the term. Large standard errors in the small survey sample at T2 (*N2T2* = 18) are likely a contributing factor to the nonsignificant results.

Table 2: Bivariate correlations of within-course centered action counts per week with self-reported self-efficacy for self-regulation.

	final grade	f20 action count	week 1 action counts	week 2 action counts	week 3 action counts	week 4 action counts	week 5 action counts	week 6 action counts	week 7 action counts	week 8 action counts	week 9 action counts	week 10 action counts
T1 SRL self-efficacy	0,13	**0,56**	**0,41**	**0,50**	**0,50**	0,15	**0,65**	**0,41**	0,34	0,16	0,26	*0,39*
T2 SRL self-efficacy	*0,43*	**0,47**	0,22	**0,48**	*0,45*	0,35	0,21	0,32	0,19	-0,17	*0,46*	0,39

Note. *NT1* = 25 students, *NT2* = 18 students. Within-course centered weekly action counts.
Greyed cells – weeks with midterm exam (week 3, 6, 9). Bolt font $p < .05$. Italic font $p < .10$.

4.3. Discussion

Descriptive and correlative findings provided in the example with combined digital trace data, course syllabus data, and survey data from the UCI-MUST project are consistent with previous literature. Variability in click activities across the weeks of the quarter and the increased number of clicks during the midterm exam weeks highlight the importance of considering course design features (i. e., exam weeks) when interpreting students' study patterns in courses. Such information can be obtained, for instance, from the course syllabi. Furthermore, our findings indicated that students' overall click activities were positively associated with final course grades. These findings corroborate previous findings (Cicchinelli et al., 2018; Nguyen et al., 2020; Park et al., 2017). Our findings further suggest that students' weekly click activities might provide valuable information on their SLR behavior. Furthermore, results indicate that click activities in specific weeks – i. e., immediately before, during, and immediately following an exam – might be particularly predictive of students' final grades.

Similar to results from prior studies (Nguyen et al., 2020; Rodriguez et al., 2021), our findings showed that students' click activities varied among students with different demographic characteristics. Female students had slightly more click activities compared to male students, whereas students from historically underrepresented minorities showed fewer click activities in their digital trace data during the weeks of the midterm exams. It is important to note that we cannot infer the causes of differing click activity patterns. Female student' higher click activities might be related to higher levels of conscientiousness. Prior SRL research has shown that more conscientious students are better in managing their time and regulating their effort in education (Douglas et al., 2016; McCrae & Löckenhoff, 2017; Waldeyer et al., 2022), and personality research has shown that female students report higher levels of conscientiousness than their male counterparts (Costa et la., 2001; Schmitt et al., 2008). Female students are also more likely to report higher test anxiety (Cassady & Johnson, 2002; Costa et la., 2001) and hence, higher click activity of female students might also be the result of increased learning activities driven by anxiety before and during exam weeks. Lower levels of click-activities among students of underrepresented minorities could be explained by other obligations, such as jobs besides studying, that might conflict with their time and resources for study activities, or by deficient SRL strategies. Further information on students' characteristics (e. g., personality traits, motivation, and goals) and their study and living situation (e. g., on other responsibilities besides studying) can be obtained through surveys, and are needed to explain variability in click activities among students. Thus, these remaining open questions indicate a need to use multiple data sources.

Our findings further showed positive associations between students' self-efficacy for SRL and their amount of click activities in the course overall, as well as in the weeks before and during exams. These findings are consistent with prior evidence of positive associations between specific SRL behaviors assessed through self-reports and digital trace data (e. g., Cicchinelli et al., 2018; Ifenthaler et al., 2022; Li et al., 2020). In our study, click activities across the quarter were positively associated with self-efficacy for SRL measures at the beginning (T1) and end of the quarter (T2), with slightly stronger associations with T1 measures. These findings are in contrast to the findings by Li et al. (2020) who reported stronger associations between self-reported data and behavioral trace data at the end of a course. These authors proposed that students rated their SRL skills based on their real experiences in the course at the end of a quarter, which likely led to more accurate self-reported SRL skills at the second time point. In our study, students were asked about their self-efficacy for SRL in general and not regarding the specific course. This might explain why associations between self-efficacy for SRL and study activities were not stronger towards the end of the quarter.

Overall, positive associations between self-report data and digital trace data on SRL and study activities point to the potential of combining both data sources to investigate study-related behaviors and academic performance: Behavioral trace data can provide measures of students' real-time study behaviors in authentic contexts. Self-report data can provide important information about a) internal cognitive and meta-cognitive aspects of SRL behavior, as well as students' self-efficacy for SRL, and b) self-report data can provide relevant information to validate new SRL measures based on digital trace data.

5. Conclusion

This contribution aimed to provide an overview of current approaches on how to investigate adaptive SRL behaviors of college students with digital trace data. Digital trace data seems particularly useful to measure SRL behaviors related to students' time management and aspects of monitoring behavior (e. g., through self-assessments with quizzes). While broad measures of study activities, such as the overall number of clicks and time spent on courses are moderate predictors of performance, more fine-grained measures, such as changes in click activities towards a deadline or using voluntary quizzes to monitor learning, are particularly promising to identify students' use of SRL strategies and associations with subsequent performance outcomes. To interpret such measures, it is necessary to take into account information about course design and context. Course syllabi and course plans can be reliable sources to obtain such relevant information (as done in the empirical example described above). Future research on SRL in college should continue to examine how information from survey data and digital trace data can be combined to investigate factors that facilitate or hinder SRL and performance in college. The above-mentioned studies successfully used digital trace data to describe adaptive and maladaptive study patterns. However, a remaining question is whether adaptive study behaviors were driven by external course designs and demands (i. e., *externally regulated* behavior), or through individual and purposefully used SRL strategies (i. e., *self-regulated* behavior). By combining digital trace data, survey data, and course syllabus data, future research could investigate the extent to which intraindividual and course contextual factors contribute to explaining variance in study behaviors and course performance. Multiple source data would allow, for instance, to investigate if regular self-testing behavior and subsequent course performance are the result of course requirements and grading policies, or of students' individual motivation and SRL skills.

References

Arizmendi, C. J., Bernacki, M. L., Raković, M., Plumley, R. D., Urban, C. J., Panter, A. T., Greene, J. A., & Gates, K. M. (2022). Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behavior Research Methods*. Advance online publication. https://doi.org/10.3758/s13428-022-01939-9

Arum, R., Eccles, J. S., Heckhausen, J., Orona, G. A., Keyserlingk, L. von, Wegemer, C. M., Wright, C. E., & Yamaguchi-Pedroza, K. (2021). A Framework for Measuring Undergraduate Learning and Growth. *Change: The Magazine of Higher Learning*, *53*(6), 51–59. https://doi.org/10.1080/00091383.2021.1987810

Baker, R., Di Xu, Park, J., Yu, R., Li, Q., Cung, B., Fischer, C., Rodriguez, F., Warschauer, M., & Smyth, P. (2020). The benefits and caveats of using clickstream data to understand student self-regulatory behaviors: opening the black box of learning processes. *International Journal of Educational Technology in Higher Education*, *17*(1). https://doi.org/10.1186/s41239-020-00187-1

Berger, F., Schreiner, C., Hagleitner, W., Jesacher-Rößler, L., Roßnagl, S., & Kraler, C. (2021). Predicting Coping With Self-Regulated Distance Learning in Times of COVID-19: Evidence From a Longitudinal Study. Frontiers in Psychology, 12, 701255. https://doi.org/10.3389/fpsyg.2021.701255

Bernacki, M. (2018). Examining the cyclical, loosely sequenced, and contingent features of self-regulated learning: Trace data and their analysis. In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 370–387). Routledge Taylor& Francis Group. https://doi.org/10.4324/9781315697048-24

Bernacki, M. L., Chavez, M. M., & Uesbeck, P. M. (2020). Predicting achievement and providing support before STEM majors begin to fail. *Computers & Education*, *158*, 103999. https://doi.org/10.1016/j.compedu.2020.103999

Boekaerts, M. (1999). Self-regulated learning: where we are today. *International Journal of Educational Research*, *31*(6), 445–457. https://doi.org/10.1016/S0883-0355(99)00014-2

Boekaerts, M. (2010). *Handbook of self-regulation*. Academic Press.

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education*, *27*, 1–13. https://doi.org/10.1016/j.iheduc.2015.04.007

Cassady, J. C., & Johnson, R. E. (2002). Cognitive Test Anxiety and Academic Performance. *Contemporary Educational Psychology*, *27*(2), 270–295. https://doi.org/10.1006/ceps.2001.1094

Cicchinelli, A., Veas, E., Pardo, A., Pammer-Schindler, V., Fessl, A., Barreiros, C., & Lindstädt, S. (2018). Finding traces of self-regulated learning in activity streams. In A. Pardo, K. Bartimote-Aufflick, G. Lynch, S. B. Shum, R. Ferguson, A. Merceron, & X. Ochoa (Eds.), *Proceedings of the 8th International Conference on Learning Analytics and Knowledge* (pp. 191–200). ACM. https://doi.org/10.1145/3170358.3170381

Cogliano, M., Bernacki, M. L., Hilpert, J. C., & Strong, C. L. (2022). A self-regulated learning analytics prediction-and-intervention design: Detecting and supporting struggling biology students. *Journal of Educational Psychology.* Advance online publication. https://doi.org/10.1037/edu0000745

Costa, P., JR., Terracciano, A., & McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, *81*(2), 322–331. https://doi.org/10.1037//0022-3514.81.2.322

Crompton, H., Bernacki, M., & Greene, J. A. (2020). Psychological foundations of emerging technologies for teaching and learning in higher education. *Current Opinion in Psychology*, *36*, 101–105. https://doi.org/10.1016/j.copsyc.2020.04.011

Dörrenbächer, L., & Perels, F. (2016). Self-regulated learning profiles in college students: Their relationship to achievement, personality, and the effectiveness of an intervention to foster self-regulated learning. *Learning and Individual Differences*, *51*, 229–241. https://doi.org/10.1016/j.lindif.2016.09.015

Douglas, H. E., Bore, M., & Munro, D. (2016). Coping with University Education: The relationships of Time Management Behaviour and Work Engagement with the Five Factor Model Aspects. *Learning and Individual Differences*, *45*, 268–274. https://doi.org/10.1016/j.lindif.2015.12.004

Farr, B., Rotermund, S., Ho, P., Radwin, D., & Robles, J. (2011). Evaluation of the Academy for College Excellence: Year 1 interim report.

Greene, J. A., Plumley, R. D., Urban, C. J., Bernacki, M. L., Gates, K. M., Hogan, K. A., Demetriou, C., & Panter, A. T. (2021). Modeling temporal self-regulatory processing in a higher education biology course. *Learning and Instruction*, *72*, 101201. https://doi.org/10.1016/j.learninstruc.2019.04.002

Hong, W., Bernacki, M. L., & Perera, H. N. (2020). A latent profile analysis of undergraduates' achievement motivations and metacognitive behaviors, and their relations to achievement in science. *Journal of Educational Psychology*, *112*(7), 1409–1430. https://doi.org/10.1037/edu0000445

Huang, X., Bernacki, M. L., Kim, D., & Hong, W. (2022). Examining the role of self-efficacy and online metacognitive monitoring behaviors in undergraduate life science education. *Learning and Instruction*, 101577. https://doi.org/10.1016/j.learninstruc.2021.101577

Ifenthaler, D., & Gibson, D. (Eds.). (2020). *Advances in Analytics for Learning and Teaching. Adoption of Data Analytics in Higher Education Learning and Teaching.* Springer International Publishing. https://doi.org/10.1007/978-3-030-47392-1

Juriševič, M., Lavrih, L., Lišić, A., Podlogar, N., & Žerak, U. (2021). Higher Education Students' Experience of Emergency Remote Teaching during the Covid-19 Pandemic in Relation to Self-Regulation and Positivity. Center for Educational Policy Studies Journal, 11(Sp.Issue). https://doi.org/10.26529/cepsj.1147

Kitsantas, A., Winsler, A., & Huie, F. (2008). Self-Regulation and Ability Predictors of Academic Success During College: A Predictive Validity Study. *Journal of Advanced Academics*, *20*(1), 42–68. https://doi.org/10.4219/jaa-2008-867

Li, Q., & Baker, R. (2018). The different relationships between engagement and outcomes across participant subgroups in Massive Open Online Courses. *Computers & Education*, *127*, 41–65. https://doi.org/10.1016/j.compedu.2018.08.005

Li, Q., Baker, R., & Warschauer, M. (2020). Using clickstream data to measure, understand, and support self-regulated learning in online courses. *The Internet and Higher Education*, *45*, 100727. https://doi.org/10.1016/j.iheduc.2020.100727

McCrae, R. R., & Löckenhoff, C. E. (2010). Self-Regulation and the Five-Factor Model of Personality Traits. In R. H. Hoyle (Ed.), *Handbook of personality and self-regulation* (pp. 145–168). Wiley-Blackwell. https://doi.org/10.1002/9781444318111.ch7

Nguyen, Q., Rienties, B., & Richardson, J. T. E. (2020). Learning analytics to uncover inequality in behavioural engagement and academic attainment in a distance learning setting. *Assessment & Evaluation in Higher Education*, *45*(4), 594–606. https://doi.org/10.1080/02602938.2019.1679088

Park, J., Denaro, K., Rodriguez, F., Smyth, P., & Warschauer, M. (2017). Detecting changes in student behavior from clickstream data. In A. Wise, P. H. Winne, G. Lynch, X. Ochoa, I. Molenaar, S. Dawson, & M. Hatala (Eds.), *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 21–30). ACM. https://doi.org/10.1145/3027385.3027430

Park, J., Rodriguez, F., Baker, R., Smyth, P., & Warschauer, M. (2018). *Understanding student procrastination via mixture models.: Paper presentation.* 11th International Conference on Educational Data Mining (EDM), Raleigh, North Carolina.

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). *A manual for the use of the motivated strategies for leaning questionnaire (MSLQ).* University of Michigan, National Center for Reserach to Improve Postsecondary Teaching and Learning.

Rodriguez, F., Lee, H. R., Rutherford, T., Fischer, C., Potma, E., & Warschauer, M. (2021). Using Clickstream Data Mining Techniques to Understand and Support First-Generation College Students in an Online Chemistry Course. In M. Scheffel, N. Dowell, S. Joksimovic, & G. Siemens (Eds.), *LAK21: 11th International Learning Analytics and Knowledge Conference* (pp. 313–322). ACM. https://doi.org/10.1145/3448139.3448169

Waldeyer, J., Dicke, T., Fleischer, J., Guo, J., Trentepohl, S., Wirth, J., & Leutner, D. (2022). A moderated mediation analysis of conscientiousness, time management strategies, effort regulation strategies, and university students' performance. *Learning and Individual Differences*, *100*, 102228. https://doi.org/10.1016/j.lindif.2022.102228

Weinstein, C. E., & Palmer, D. R. (2002). *Learning and Study Strategies Inventory (LASSI): User's manual* (2nd ed.). H&H Publishing.

Wolters, C. A., & Won, S. (2018). Validity and the use of self-report questionnaires to assess self-regulated learning. In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 304–322). Routledge Taylor& Francis Group. https://doi.org/10.4324/9781315697048

Zimmerman, B. J. (1990). Self-Regulated Learning and Academic Achievement: An Overview. *Educational Psychologist*, *25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2
Zimmerman, B. J. (2000). Attaining Self-Regulation. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7
Zimmerman, B. J., & Schunk, D. H. (Eds.). (2011). *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (1. publ). Routledge. https://doi.org/10.4324/9780203839010