



Analyzing productive learning behaviors for students using immediate corrective feedback in a blended learning environment

Xin Chen^{a,*}, Lori Breslow^b, Jennifer DeBoer^a

^a School of Engineering Education, Purdue University, 701 W. Stadium Ave., Suite 1300, West Lafayette, IN 47907, United States

^b MIT Sloan School of Management, Massachusetts Institute of Technology, 77 Massachusetts Ave., Building 62–339, Cambridge, MA 02139, United States

ARTICLE INFO

Keywords:

Computer-mediated feedback
Blended learning
Human-computer interface
Physics education

ABSTRACT

Undergraduate classes in many science and engineering courses are utilizing asynchronous computer platforms to host educational materials such as lecture videos or discussion forums. These platforms also have the ability to provide immediate feedback to students on formative assessment tasks such as homework problems, reading questions, or weekly quizzes. Although there have been a number of studies on computer-based feedback, there is more we need to know about how students interact with immediate feedback, and how those interactions influence their learning. In this study, we characterize introductory physics students' interactions with one computer-based immediate simple corrective feedback tool, the “checkable answer feature” (CAF), powered by the institutional version of the edX platform. We investigate how much students interact with the CAF, the patterns of interaction, and, ultimately, how these patterns are associated with course performance. We utilize rich quantitative data, including a large volume of server tracking logs that show students' use the CAF, as well as performance metrics. Our findings show certain patterns of engagement with feedback reflect productive study strategies and significantly predict higher performance. The findings provide guidance for instructional practice and the continued development of online feedback tools in introductory STEM courses.

1. Introduction

In educational settings, feedback serves the crucial role of closing the gap between students' current understanding and desired learning (Shute, 2008). Timely and informative feedback can help learners recognize and correct misconceptions, motivate them to acquire knowledge, and increase their confidence and motivation to learn (Epstein et al., 2010). Yet in the context of higher education, it is often not possible for instructors to provide timely feedback to every student individually. This is especially true in first-year foundational courses due to the large number of students usually enrolled. But feedback delivered on computer-based platforms such as edX (www.edx.org), Coursera (www.coursera.org), and FutureLearn (www.futurelearn.com) offer a potential solution. They can provide students with automatic immediate feedback on formative assessment tasks such as reading questions, homework problems, and short quizzes. These platforms enable the detailed recording of the activities the students perform online. In turn, these data logs can be re-assembled to give researchers an in-depth look into how students are learning the concepts and skills instructors want them to master.

The specific feedback mechanism we focus on in this study is the “checkable answer feature” (CAF) included in a suite of online

* Corresponding author.

E-mail addresses: cynthiachen.ee@gmail.com, cindy.chen@airbnb.com (X. Chen), lrb@mit.edu (L. Breslow), deboerj@purdue.edu (J. DeBoer).

ROTATING AND TRANSLATING BOWLING BALL BACKSPIN (10 points possible)

This problem must be submitted online.

A bowling ball of mass m and radius R is initially thrown down an alley with an initial speed v_0 and backspin with angular speed ω_0 , such that $v_0 > R\omega_0$. The moment of inertia of the ball about its center of mass is $I_{\text{cm}} = (2/5)mR^2$. What is the speed v_f of the bowling ball when it just starts to roll without slipping?



SYMBOLIC CHECK

For the symbolic check, express your answer using some or all of the following variables: R , v_0 for v_0 , ω_0 for ω_0 .

$v_f =$ ❌

Fig. 1. Screenshot of the checkable answer for an online homework problem.

materials in an introductory blended learning course in physics, which we call PHYS101, at an elite private university in the northeast United States. The CAF, a rudimentary intelligent tutoring system, gives students immediate feedback on whether or not their answers to online homework problems are correct. Once the student enters their answer, if it is correct, the system returns a green tick mark, and if it is wrong, they see a red X (see Fig. 1). In addition, students use the CAF to see if their answers to multi-part problems, which they have to write out, are correct before they submit their traditional, paper-based homework assignments.

It is evident that formative feedback can foster improved achievement and enhanced motivation to learn, while also supporting deep and self-regulated learning (Crisp & Ward, 2008; Koh, 2008; Wolsey, 2008). Studies of courses that are wholly online (Crisp & Ward, 2008; Gijbels, Dochy, Van den Bossche, & Segers, 2005; Sorensen & Takle, 2005; Van der Pol, Van den Berg, Admiraal, & Simons, 2008; Vonderwell, Liang, & Alderman, 2007; Wang, Wang, & Huang, 2008) show that the effective use of online formative assessment can engage students and teachers in meaningful educational experiences, providing them with opportunities to collaboratively identify learning needs and devise strategies to meet them. Other studies have investigated the effectiveness of online or computer-assisted feedback tools in blended learning environments (e.g., Angus & Watson, 2009; Chung, Shel, & Kaiser, 2006; Feldman & Capobianco, 2008; Lin, 2008; Wang et al., 2008). Foundational work by Van der Kleij, Feskens, and Eggen (2015) analyzing a significant number of studies indicates that elaborated feedback is more effective than simple corrective feedback. This effect is more pronounced in higher level learning and in certain disciplines, as, for example, math versus science. Interestingly, several studies, including recent work by Van der Kleij, do not observe a difference between corrective feedback and elaborated feedback (Mazingo, 2006; Van der Kleij, Eggen, Timmers, & Veldkamp, 2012).

This raises the possibility that how students engage with feedback may have a significant influence on how feedback affects learning outcomes. While, there is evidence that the amount of time spent with feedback is mediated by student attitudes and motivations (Narciss & Huth, 2006; Van der Kleij et al., 2012), to our knowledge there are no quantitative studies of how the level and patterns of student engagement with corrective feedback can affect learning outcomes. Our study addresses this gap by addressing the following research questions:

RQ1: How much do students interact with simple corrective feedback in a blended undergraduate physics class, and what are the patterns of engagement?

RQ2: (a) What patterns of engagement with simple corrective feedback comprise productive study behaviors and are associated with stronger performance, and (b) what patterns are negatively correlated with performance?

Computer-based feedback systems are becoming more sophisticated and of growing benefit in higher education. Individual feedback can be more easily scaled to large numbers of learners because each student can interact with information that is instantaneously provided, and the material can be focused on the individual student's cognitive needs, study strategies, and preferences for how material is presented. By combining advances in artificial intelligence with content experts' knowledge of difficult concepts and common misconceptions, the material can also address the stumbling blocks that interfere with learning in particular disciplines. Feedback can be delivered when and where it is most convenient for the student (a dorm room, on-campus common study space, or Starbucks, for that matter). The ever-increasing use of these online educational platforms, coupled with rich, large-scale data gathering tools, enables an unprecedented window into student use of computer-based feedback mechanisms. Recognizing these

interactions can help instructors deliver more targeted interventions, as well as help platform developers create the features that can help maximize the potential of these online systems. This study also contributes to understanding how a large, detailed online dataset can reveal patterns of student study behaviors. In turn, this demonstrates ways in which data mining can be used in to complement traditional statistics and generate new knowledge. Finally, correlating online feedback behaviors with course performance sheds light on how self-regulated learning and productive struggle contribute to positive student outcomes.

2. Context of the study

Our specific study was undertaken within the context of a semester-long required residential physics course in mechanics, which all undergraduates at the university must pass to graduate. A well-respected faculty member in the physics department began to experiment with redesigning the course in 2000, moving away from lectures to a model that incorporated active learning pedagogies. (In physics, this model of teaching is known as “studio physics” (Beichner et al., 2007; Beichner & Saul, 2003; Wilson, 1994).) The transformed course was launched at full scale in 2003. Students take the course in one of two specially built classrooms, sitting at round tables of nine and working in groups of three. Significant technology was built into the room, including white boards and AV screens along the entire perimeter of the room. Each table has a video camera trained on it so students can present their work to the entire class. The instructor’s “command center” was consciously placed about a third of the way from the front of the room to signal that they were not the sole authority in the classroom (Breslow, 2010).

Many materials were put onto the university’s learning management system to accompany this new pedagogical model. But it is important to note that students were still learning the material in face-to-face classrooms, as they continue to. Classroom time entails two, 2-h classes each week, during which instructors present short lectures and then students *do* something with the concept just explained. That activity might be to carry out a desk-top experiment, analyze data, or solve a problem similar to those on the homework assignment for that week. Every Friday, the class meets for an hour and students solve problems in teams with instructors and teaching assistants available in the room.

The next step in strengthening the course occurred in 2014, when many of the original instructional team collaborated with developers to put all the course materials on the university’s local version of the open edX massive open online course [MOOC] platform. The instructional team took advantage of this migration to put more resources online, including lecture videos, animations and simulations, laboratories (either interactive or static), recordings of an instructor working through problems in a tutorial setting, PDFs of notes pertaining to the homework assignment for the week, and a commercially based discussion forum. At the same time, the CAF was built and incorporated into the online course materials.

In fall semester 2014, when our study took place, the total enrollment in PHYS101 was 474 students. The course lasted for 15 weeks, and the students were assigned one or two online homework problems during each of the first 12 weeks. In total, there were 22 online homework problems throughout the semester. For the most part, the online homework problems asked students to solve a classical mechanics problem, usually to the point of an algebraic solution.

The students had 10 days to work on each problem before the due date. They submitted their answers online, and when they did, the platform would tell them immediately if they were right or not by returning either a green tick mark or a red X. Students could check an unlimited number of times whether their answers were correct, though no answer or hint was available before the due date. Additionally, students had to submit written homework each week that consisted of five to six problems for which they had to write out the numerous steps they took to solve the problem. They could use the CAF to check interim answers to these multi-part problems. The scores for the online homework counted for 2% towards the cumulative grade in the course, and the written problems sets were worth a total of 8%. Students also received points for completing short assignments based on the readings to be done before class, as well as for participation in in-class activities. The remainder of the cumulative grade was based on three exams (15% each) and a final (25%). The results of a survey about the CAF administered in 2014 showed overwhelmingly that the students liked it, and they recommended that CAFs be implemented in other physics courses (Rayyan & Belcher, 2014).

By 2014, then, the course was moving to a more blended model as students had a number of online resources they could access. However, we recognize that the definition of “blended learning” is still a matter of debate in the field. If we rely on Allen, Seaman and Garrett’s definition (2007), a learning environment is blended when 30–80% of the learning/teaching activities are conducted through web-based tools. Because so much of what the students were learning was through online resources, we argue, as does the lead course instructor, that by 2014, PHYS101 was a blended course (S. Rayyan, personal communication, September 15, 2017). We also acknowledge that instructors and researchers have still not reached consensus about how to best marry the increasing use of and dependence on technology with face-to-face class time, and that our study takes place within this ill-defined context. There are virtually countless possibilities for instructional and platform design online, depending on instructional goals and resources, disciplinary variables, and differences in student backgrounds and capabilities in a blended learning context (Bernard et al., 2009; Bernard, Borokhovski, Schmid, Tamim, & Abrami, 2014; Lack, 2013; Zhao & Breslow, 2013). As Bernard and his colleagues (2014) write, and we concur, studies that answer the big questions (e.g., is blended learning more effective than traditional classroom instruction?) “generally fail to establish an alignment of evidence that addresses the ‘do’s’ and ‘don’ts’ of instruction via rigorous research” (p. 89). Our study does not intend to determine whether blended learning is effective or not. Rather, we hope to add to an understanding of how student engagement with the affordances of online feedback plays a role in the learning process of STEM undergraduates.

3. Theoretical framework

This study is grounded both in the research on feedback in learning, which has a long and rich history, and relatively new scholarship on computer-mediated feedback, including comparisons between students' use of feedback provided by intelligent tutoring systems and online learning platforms. Some of the behaviors we see indicate a subset of the students are using strategies associated with self-regulation, and we utilize that body of work as well. We knit this applicable background literature into a theoretical framework to guide our research.

3.1. *The role of feedback in learning*

Even within a blended learning environment, one could argue that in-person feedback provided by instructors continues to be the best source of feedback. Based on social constructivism learning theory (Papert, 1993; Vygotsky, 1986) and numerous theoretical and empirical studies on how people learn (e.g., Bransford, Brown, & Cocking, 1999), we know that students begin to form their understanding of a concept from their prior knowledge and experiences. But feedback, defined as the “information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding” (Hattie & Timperley, 2007, p. 81), plays an important role in closing the gap between current understanding and the desired learning goal. If a concept is faulty in the student's mind and they rely on that concept as a foundation for further understanding, they are unlikely to make much progress in learning. In fact, holding onto a misunderstanding can potentially negatively impact mastery (Pellegrino, Chudowsky, Glaser, & National Research Council, 2001).

To understand the effect of feedback on learning, we need to first understand its numerous dimensions. The two most rigorously studied characteristics of feedback are types of feedback and when students receive it. Hattie and Timperley (2007) did a thorough review of studies on feedback effects, synthesizing four major levels of feedback from specific to generic: (1) feedback on task execution, (2) feedback on learning strategies, (3) feedback on metacognitive skills, and (4) feedback on the self as a person. Researchers posit that feedback about learning strategies and metacognitive skills is more effective at enhancing deeper learning than the kind of surface knowledge gained from feedback on task execution (Balzer, Doherty, & O'Connor, 1989; Earley, Northcraft, Lee, & Lituchy, 1990). However, more complex feedback usually imposes a heavier cognitive load on students. (It is also more challenging to build into a computer-mediated system, and thus it is less scalable.) There is no reason that the student cannot be given different levels of feedback at the same time, and, in fact, a positive interactive effect between feedback on task execution and on learning strategies has been found (Earley et al., 1990). Feedback at the task level can help students who already demonstrate metacognitive awareness to develop appropriate learning strategies on their own, which can be more powerful than direct feedback about the strategies. This is to say, although the checkable answer feature in the course we study only provides simple corrective feedback at the task execution level, it is reasonable to assume that students with strong metacognitive awareness could develop effective learning strategies from that feedback. For instance, they could gain skills in diagnosing errors or even skills such as self-assessment and self-reflection. These students may then demonstrate highly strategic and productive behaviors that benefit their learning.

The studies on the timing of feedback tend to focus on immediate versus delayed feedback, with much debate on the effectiveness of one in comparison to the other (e.g., Kulhavy & Anderson, 1972; Kulik & Kulik, 1988). Although delayed feedback appears to be more effective in laboratory studies, applied studies in the classroom usually produce the opposite results. Potentially, this is because immediate feedback reduces preservation of initial misconceptions and also increases students' confidence and motivation to learn (Dihoff, Brosvic, & Epstein, 2003; Dihoff, Brosvic, Epstein, & Cook, 2004). Again, our study focuses on immediate corrective feedback at the task level, but in our discussion section, we briefly explore potential implications for the timing and levels of feedback that we can infer from our results.

3.2. *Computer-mediated feedback*

Computer-mediated feedback provides feedback to learners in an efficient, scalable, and trackable way. A recent foundational meta-analysis that explored a large number of studies of item-based feedback estimated larger positive effects for elaborated feedback than correctness feedback (Van der Kleij et al., 2015). In particular, elaborated feedback was more helpful for higher level learning outcomes. In our study, however, we focus on simple correctness feedback and introductory level learning outcomes, which Van der Kleij et al. (2015) show have a small positive relationship. Our study can nuance these results by differentiating how users employ different strategies for engaging with correctness feedback, which ultimately relate to differential learning outcomes. Interestingly, another study comparing correctness feedback to elaborated feedback found that, while elaborated feedback is helpful, it is more dependent on students' attitudes and motivation. However, correctness feedback was consistently better than no feedback at all (Maier, Wolf, & Randler, 2016).

Often, the literature on computer-mediated feedback compares the effectiveness of different levels of feedback, with limited distinction regarding the *amount* of engagement with feedback (e.g., how many times do they check for their results?) and the *patterns* of engagement (e.g., how spaced out are their checks?). A few studies have suggestive findings that motivate our study of how students engage with correctness feedback. Research undertaken in face-to-face classrooms has found that implementing an answer-until-correct requirement and providing immediate feedback yields gains in learning (Clariana, 1990; Persky & Pollack, 2008). In an online environment, Narciss and Huth (2006) show that displaying the correct answer only when the student gets the question right, which is what the CAF does, has positive motivational effects. Similarly, frequent simple multiple-choice questions with correctness feedback during weekly lectures were shown to improve performance and reduce achievement gaps (Pennebaker, Gosling, & Ferrell,

2013). We add to this literature by quantitatively describing students' behaviors in engaging with one promising, scalable type of feedback in an introductory science course.

In order to do so, we need to determine how to operationalize the “amount” of engagement, as well as what behaviors we can observe to describe “patterns” of engagement. Given the limited amount of literature on engagement patterns using computer-mediated feedback, we drew on a broader set of behaviors observed in computer-mediated higher education. Wei, Peng, and Chou (2015) study the records of how students used online tools in a blended learning course, and they apply a framework that we find useful to motivate the different types of behaviors we measure. They study interactions between: learner-self, learner-interface, learner-content, learner-instructor/learner-learner. We can apply this framework to our own study by looking at patterns of learner-content activity, for example, student interactions with course videos. Wei et al. (2015) further delineate the amount of engagement they see by capturing absolute numbers of interactions, frequencies of logins and interactions, and, implicitly, the durations of interactions and student study sessions. We extend their metrics to the checkable answer feature and the numerous types of content covered in the course we study. Zheng and Warschauer (2015) corroborate these operationalizations of student engagement in a different context—a study of student participation and interaction in a solely online discussion environment. Their hierarchical linear model measures student engagement again with frequencies of posting and variation in posting over time. Their study also highlights the need to control for student demographics and prior educational outcomes, which we do.

Finally, we know from the literature on intelligent tutoring systems (ITS) and massive open online courses (MOOC) that some students use incorrect answers to cheat or to get the correct answers more easily. This is known as “gaming” behaviors (Alevén & Koedinger, 2000; Baker, Corbett, Koedinger, & Wagner, 2004; Northcutt, Ho, & Chuang, 2016; Ruiperez-Valiente, Alexandron, Chen, & Pritchard, 2016). Although ITSs and online learning platforms are usually used in different contexts, literature on student behaviors using ITSs nevertheless reminds us that students in online learning environments could demonstrate a variety of behaviors when using the immediate feedback features, some of which are unintended by the instructors and platform designers. These different behaviors might either benefit or hurt student learning.

3.3. Study strategies

As more and more instruction is moved online, students will need to accept additional responsibility for their own learning. It is important to understand how the beneficial approaches and activities students undertake can be promoted in blended or online environments. For example, He, Gajski, Farkas, and Warschauer (2015) found metacognitive skills such as time management, goal setting, and self-monitoring to be key mediators in a blended learning environment.

Many of these advantageous behaviors have been identified by studies in more traditional educational settings. For example, a rigorous body of research exists on self-regulated learning (SRL), a complex construct in educational psychology (e.g., Boekaerts, 1999; Pintrich & de Groot, 1990; Zimmerman, 1990). SRL connects the goals, motivations, and desires of the individual student to the complex system in which learning takes place. As all complex systems, learning is composed of many interacting parts that affect one another to create specific learning outcomes for individual learners.

Students who exhibit SRL are often more motivated by the goals they have for their own learning than by extrinsic rewards. SRL encompasses metacognitive processes, as listed above, but also includes elements of motivation, including the belief in one's own capabilities to master a set of concepts or skills (Pintrich & de Groot, 1990). The behaviors self-regulated learners exhibit include setting goals for themselves, monitoring their progress, and creating environments that foster their learning. These students realize they have control over their learning, recognize what they know and do not know, and accept responsibility for addressing their gaps in knowledge. They understand that learning is a process they can control, and they actively participate in that process rather than passively responding to instructions from their teachers. They create an environment around them that will support their learning. They may take the initiative to ask their instructors for additional help, they may set up a physical space that is conducive to learning, and/or they may reflect on what they have learned and use a feedback loop to further mastery. They realize how and when they have been successful and continue to use those methods.

A similar way of looking at these kinds of attitudes and behaviors is captured by the phrase “productive struggle.” First coined by Uri Treisman, “productive struggle” also describes a set of skills students use to further their learning, as well as a tenacity they manifest (as quoted in Silva & White, 2013, p. 9). Researchers at the Carnegie Foundation for the Advancement of Teaching have adopted the phrase “productive struggle” to describe the process where students work through difficulties but eventually reach their learning goals. Students who engage in productive struggle realize that learning a new idea or skill is often hard work, but they do not abandon the effort; they persevere. Vedel (2014) suggests that one of the Big Five personality traits, conscientiousness, may also reflect these attitudes and set of behaviors.

In recent studies, Krumm and colleagues develop measures for students' “productive persistence,” or strategic behaviors that benefit learning in an online learning platform. To operationalize this construct, they use the percent of on-line sessions wherein students attempted another activity after they gave an incorrect answer to an online problem. In our study, we investigate whether there are distinguishable groups of students who demonstrate self-regulated learning and productive study strategies using the CAF, and if those behaviors correlate with stronger course outcomes.

4. Methodology

In order to identify students' behavioral patterns with the CAF and to understand the relationship between those patterns and performance metrics, we needed to go through several distinct steps, balancing theory with emergent patterns we saw in our data.

Table 1
Demographic variables of students in PHYS101.

Demographic	% of Total
<i>Self-reported gender</i>	
Male	46.5
Female	53.5
<i>Race/ethnicity</i>	
Caucasian	39.5
African American	10.4
Asian	26.1
Hispanic	16.1
Other	7.9
<i>Pre-college bridge program participation</i>	
Yes	8.0
No	92.0
<i>Parental education</i>	
First Generation Students	20.0
Non-First Generation Students	80.0
<i>Socio-economic status</i>	
Tuition-Free Students	21.1
Non-Tuition Free Students	78.9

First, we generated descriptive statistics, and then we used inferential statistics and data mining techniques to determine the relationship between student behaviors and their outcomes. This approach allows us to take advantage of traditional methods in educational research, as well as techniques being used to analyze the large amount of data collected by learning management systems, MOOC platforms, and intelligent tutoring systems.

4.1. Sources of data

We have three sources of data for this study. The university provided us with demographic data that gave us information about the students' backgrounds. But for our study's largest and most novel source of data, we had access to the detailed tracking logs that recorded every time the students clicked on the course website. We were also given course performance metrics, including scores on the final exams and cumulative grades in the course. In addition, we administered a survey that asked students about their self-efficacy towards the course content, their perceptions of the relevance for the material in the course, and their perceptions and self-reported use of the CAF, but the results of the survey will be reported in another paper. Below we provide more detail on the data sources for this study.

4.1.1. Student demographics

Data on student background factors were provided by the university's Office of Institutional Research, the Office of the Dean of Admissions and Financial Aid Services, and the Registrar. Demographic information such as gender, ethnicity, family income, and parental educational attainment are detailed in Table 1. We include these demographics as control variables in our regression model, with the additional controls of prior educational ability and experience (pre-college bridge program participation, prior or concurrent math courses, and SAT II science scores).

4.1.2. Tracking logs

From September 1 to December 31, 2014, students' interactions with the course website resulted in more than 30 million browser-side and server-side interactive events, which constituted more than 8 GB of raw log data. These events included students' interactions with numerous kinds of online materials and assessment tasks, including an e-textbook, lecture slides, PDF notes, videos, reading summaries, and homework problems. For the CAF, specifically, one interaction event is referred to as a "problem-check" or simply a "check." In total, students made 58,422 checks while completing the 22 online homework problems. The total number of correct checks made for online homework problems was 9,545, and that of incorrect checks was 48,877, which means that only 16.3% of the checks were correct.

4.1.3. Performance metrics

For all students enrolled in this course, we collected the final exam scores and the cumulative grades in the course, as detailed in Table 2. As many assignments and activities contributed to the cumulative grade, it represented students' effort to a larger extent than the final exam grades did.

4.2. Analysis

As noted above, we undertook the analysis of the data in several steps. First, we identified appropriate measures for the ways students interacted with the CAF and the other online resources and generated overall descriptive statistics. We then applied a series

Table 2

Final exam scores and cumulative grades for students in PHYS101.

	Mean	Median	Standard Deviation	Min	Max
Final Exam	137.7	139	29.3	0	196
Cumulative Grade	79.9%	80.3%	8.6%	21.7%	97.2%

of statistical and data mining methods, including correlation analysis, agglomerative hierarchical clustering, K-means clustering, and multiple linear regression with LASSO (Least Absolute Shrinkage and Selection Operator) variable selection to discern patterns of students' behaviors. As we did these statistical tests, we also referred to the literature to help us make sense of the patterns we were seeing inductively. Finally, we ran t-tests and multiple linear regression to determine the relationship of students' behaviors with their final exam and cumulative grade scores in the course.

4.2.1. Identifying behavior variables: reciprocity between literature and patterns in the data

Again, the process of generating variables was informed by a back and forth between inductive and deductive methods. We looked for patterns in the data to describe all of the student behaviors we could discern, and then chose the variables we did informed by our theoretical framework and previous studies. Please see Table 3 for a complete list of the 30 variables we identified, and their range, mean, and median values. The unit we use to report time varies in order to make the time values more readable (e.g., minutes, hours vs. days).

In order to determine the frequency and duration of behaviors with the CAF, we first had to define a *problem-solving session*. Students submitted answers to each of 22 homework problems: If the answer was right, we coded it as True and if it was wrong, it was coded False. (For example, if a student submitted an answer five times for one problem, and only the last answer was correct, then this behavior was indicated using the sequence [F, F, F, F, T].) Different students had different numbers of Fs and Ts on each homework problem with different time intervals in between. If any two adjacent answers within a successive set of answers had less than a 1-h time interval in between them, we treated this as a dense cluster. The time period starting at 30 min before the first submission and ending at 30 min after the last submission in this dense cluster constituted a *problem-solving session*.

Once a problem-solving session was determined, we could identify various behaviors from the tracking logs as the students engaged with the online problems. For example, we began to see patterns in the frequency and duration with which students worked on the problems, the relationships between correct and incorrect checks, and the online resources students accessed while solving the problems.

Simultaneously, we looked at the literature to help us recognize these patterns. For instance, the research on feedback helped us identify what became variables 11, 12, and 20: *the average time between the first check and the second check for each problem; the average interval between all checks; and the average interval between checks within problem-solving sessions*, respectively. However, research on feedback also raised questions in our minds about how the students were using the red X's and green tick marks they got from the CAF. If a student's number of checks was very high, and the intervals between checks were very short, that could mean they were demonstrating relatively more hasty, superficial, or careless checking behaviors. Or, the student could be using the simple corrective feedback from the CAF more strategically to improve how they learned mechanics and to strengthen their metacognitive skills. We searched the data to try to discern this difference.

Another example of how we used previous research comes from a study about student participation in an online course. Morris, Finnegan, and Wu (2005) conclude that engagement metrics—measured both in frequency of using online materials and duration of participation—predict nearly one third of the variability in achievement in an online course. Variables 21 to 30 (e.g., *average number of video clicks within each session; average time spent checking the previous exam materials within each session*) describe the frequency and duration with which students used various online resources, and we posit these would explain a notable portion of the variance in achievement. A final connection to the literature comes from the work on self-regulation and productive struggle, as described above. This literature led us to identify variables such as *activity after incorrect* (variable 15), which captures student engagement with resources potentially used as help-seeking references after initially getting a problem wrong.

We also wanted to take into account differences in how hard the problems were, and the data from the tracking logs allowed us to estimate this. For some problems, most students could get to the correct answer within a few checks, while other problems took the same group of students tens or even hundreds of checks. We used the total number of incorrect checks for each homework problem across the entire number of students in this study to approximate the difficulty level of a given homework problem. For variable 5 (*fraction of attempted problems where the student got it correct in their first try weighted by the difficulty level of the problem*) and variable 6 (*fraction of attempted problems where the student got it correct in their last try weighted by the difficulty level of the problem*), whether students got the correct answers on their first try or last try were initially binary descriptors of either 1 or 0 for each problem. But then we weighted these variables using the total number of incorrect checks for each homework problem across all students to identify the problems as more or less difficult. If, for example, we were comparing problem A to problem B, and problem B had more incorrect checks, whether students were cheating or randomly guessing on problem B, we assume that it is more difficult.

4.2.2. Determining relationships between behaviors

After identifying the 30 behavior variables, we performed two types of analyses to estimate the relationship between behaviors. First, we conducted bivariate Spearman correlation analysis between each of the behavioral variables. We chose Spearman rank over

Table 3
Descriptive statistics for all variables.

	Description	Mean	Median	Min	Max
1 Correct Checks	The average number of correct checks for each problem.	0.92	0.95	0.05	1.77
2 Incorrect Checks	The average number of incorrect checks for each problem.	4.69	4.00	0.23	36.73
3 Total Checks	The average of the total number of checks for each problem.	5.60	4.95	0.95	37.32
4 Correct Fraction	The average fraction of correct checks for each attempted problem.	0.41	0.39	0.13	0.87
5 First Correct Fraction (before weighting using difficulty levels)	The fraction of attempted problems where the student got it correct in their first try weighted by the difficulty level of the problem.	0.19	0.16	0.00	0.69
6 Last Correct Fraction (before weighting using difficulty levels)	The fraction of attempted problems where the student got it correct in their last try weighted by the difficulty level of the problem.	0.83	0.88	0.00	0.90
7 Not Attempted	Number of not attempted problems.	1.16	0	0	17
8 First to Due	The average time between the first check and the due time for each problem.	69 h	64 h	5 h	9 days
9 Last to Due	The average time between the last check and the due time for each problem.	58 h	53 h	3 h	8 days
10 First to Last	The average time between the first check and the last check for each problem.	13 h	11 h	4 min	61 h
11 First to Second	The average time between the first check and the second check for each problem.	3 h	1 h	22 s	20 h
12 Interval between Checks	Interval between checks: The average time interval between all checks for each.	3 h	2 h	90 s	19 h
13 Overlap Time	The average overlapping time between the two online problems for each week where the student has worked back and forth between the two HW problems (overlap).	12 h	8 h	107 s	4 days
14 Weeks Overlap Exists	The number of weeks where there was overlapping between the two online problems.	5.78	6	0	9
15 Activity After Incorrect	The fraction of incorrect checks where other activities appear after the incorrect checks.	0.03	0.02	0.00	0.18
16 Num Session	Number of problem-solving sessions.	20.98	20	5	49
17 Time All Sessions	The total length of all problem-solving sessions.	27 h	26 h	6 h	59 h
18 Avg Session Length	The average length of all problem-solving sessions.	77 min	76 min	62 min	107 min
19 Interval between Sessions	The average time interval between problem-solving sessions.	4 days	4 days	2 days	13 days
20 Interval within Sessions	The average time interval between checks within each problem-solving session.	4 min	4 min	21 s	13 min
21 Video Time	The average time spent viewing videos within each session. Based on industry accepted norm for timeout (Google, 2015; Interactive Advertising Bureau, 2009), if two successive activities appear within less than 30 min, then this time period in between the two activities is counted as spent on the first activity. The following variable 23, 25, 27 were also calculated based on this practice.	18 min	0	0	3 h
22 Video Num	The average number of video clicks within each session.	16.62	0	0	371
23 Text Time	The average time spent browsing the e-textbook within each session.	49 min	29 min	0	6 h
24 Text Num	The average number of times that the student accessed the e-textbook within each session.	94.54	45	0	927
25 ClassP Time	The average time spent checking the online PDF notes for in-class problems within each session.	18 min	6 min	0	4 h
26 ClassP Num	The average number of times that the student accessed the PDF notes for in-class problems within each session.	3.62	2	0	62
27 FridayP Time	The average time spent checking the Friday problem-solving session PDF notes within each session.	7 min	0	0	2 h
28 FridayP Num	The average number of times that the student accessed the Friday problem-solving session PDF notes within each session.	1.27	0	0	20
29 Exam Time	The average time spent checking the previous exam materials within each session	7 min	0	0	3 h
30 Exam Num	The average number of times that the student accessed the previous exam materials within each session.	0.84	0	0	26

Pearson correlation because, as shown in Fig. 2, the distributions of most variables are highly skewed. (Interestingly, although these behaviors demonstrate a high level of skew, they are not as notably skewed or widely varied as behaviors in the solely online MOOC context (DeBoer et al., 2013).)

We then performed agglomerative hierarchical clustering, which, the name implies, is a bottom-up method (Rokach, 2010). Each cluster, containing only one variable at the beginning of the process, is merged with another cluster having the shortest distance (ave. Spearman correlation) as the hierarchy moves up. This clustering algorithm runs recursively until all the variables are grouped together. Agglomerative hierarchical clustering allowed us to identify variables that stem from the same latent trait. If we included both in our regression, the variance of our estimates would be inflated. It also permitted us to determine what variables were acting together as “families” of behaviors, to better understand the complex patterns students displayed in engaging with the online feedback. Again, the resulting variable groups aided us in addressing variance inflation in our traditional regression model.

4.2.3. Using variables to explain performance

To understand the relationship between each of the behavioral variables and course performance, we first calculated the correlation between each variable and the students' performance metrics (Appendix A). Next, to understand the amount of variance in students' course performance explained by students' behaviors, we performed multiple linear regression analysis. All behavioral variables were normalized to values between 0 and 1. We included important student background factors (i.e., demographics and prior ability and coursework, as described above) in the analysis.

Since many of the variables were correlated with each other, we had a low level of tolerance for many of the predictors. When we included all behavioral variables in the regression model, the variance inflation factors (VIF) for these variables ranged from 1.26 to 612.74. (VIF = 1 if none of the variance in that variable is shared with other regressors.) This could be because many variables were likely referring to closely related behaviors or mediated by similar background factors. Therefore, for each pair of variables with $|\text{bivariate correlation coefficient}| > 0.85$, we removed one of the variables from that pair.

We then used the LASSO (Least Absolute Shrinkage and Selection Operator) method for variable selection, aiming for the lowest number of variables that comprise a fully specified model and explain a similar amount of variance in the dependent variable. In other words, we balanced our need for a fully specified model that captures theoretically relevant predictors of performance with the recognition that some of the variables we extracted may be representing nearly the same behaviors. Appendix A contains the final list of behavioral variables we selected included: *weeks overlap exists*; *last correct fraction*; *avg. session length*; *first to due*; *incorrect checks*; *overlap time*; *exam time*; *video time*; *activity after incorrect*; *text time*; *first correct fraction*.

After applying the LASSO variable selection, we performed F-tests to compare the regression models before and after the variable reduction to test whether the reduction of variables significantly changed the amount of variance explained in the students' performance metrics ($F = 0.77$, $p = 0.75$). We used the residual plot and Cook's distance to identify and remove outliers and the Shapiro-Wilk test to verify that the residuals were normally distributed. Finally, OLS multiple linear regression was run using the chosen behavioral variables and control variables with the dependent variable as either the cumulative grade or the final exam grade.

As the last step in the process of identifying online behaviors that lead to stronger performance, we used the K-means clustering method to identify “types” of students who demonstrate different behavioral patterns. Our unit of analysis here is the student, each with 30 attributes (In comparison, when analyzing the hierarchy of relationships among behaviors, the unit of analysis was the behavioral variable, each with 474 values from all students.) Our goal was to identify the various student clusters and thereby compare their performance metrics. That is, we transposed the input student-variable matrices. We tested for 2, 3, 4, and 5 clusters and found two student clusters with distinctive behaviors.

5. Results

The results of our analyses allow us to begin to answer the research questions that motivate this study. We can identify how students are using immediate feedback in solving online homework problems and what behaviors correlate with a stronger performance in the course.

5.1. RQ 1: How much do students interact with simple corrective feedback in a blended undergraduate physics class, and what are the patterns of engagement?

Students are indeed utilizing the CAF as a resource and are checking each problem, on average, about five times. A few outliers demonstrate the extremes of this behavior—some students check a single problem as many as 300 times, and some do not check various problems at all—but the bulk of the students are exhibiting an “in-between” level of engagement. Interestingly, students looked at other resources after getting an incorrect check only 3% of the time. During any given problem solving session, students spent an average of 49 min browsing the e-textbook, 18 min watching videos, 8 min looking at PDF notes of in-class problems, and 7 min looking at problem-solving session notes.

On average, students spent about 27 h engaged in online homework problem solving sessions throughout the semester. Given the 22 online problems over the semester, this averages out, roughly, to about 74 min on each problem. The maximum number of checks over all 22 problems for a single student is 821, while the minimum number is 21. The average number of checks for a student is 123 and the median number is 109. Fig. 3 is a stacked bar graph illustrating the number of correct and incorrect checks for each student (ordered based on total number of checks descending). For problems where the students worked until correct, 97.3% only had 1 correct check, a near-constant. The percent of students who attempted 20 or more homeworks is 87.6%, also a near-constant.

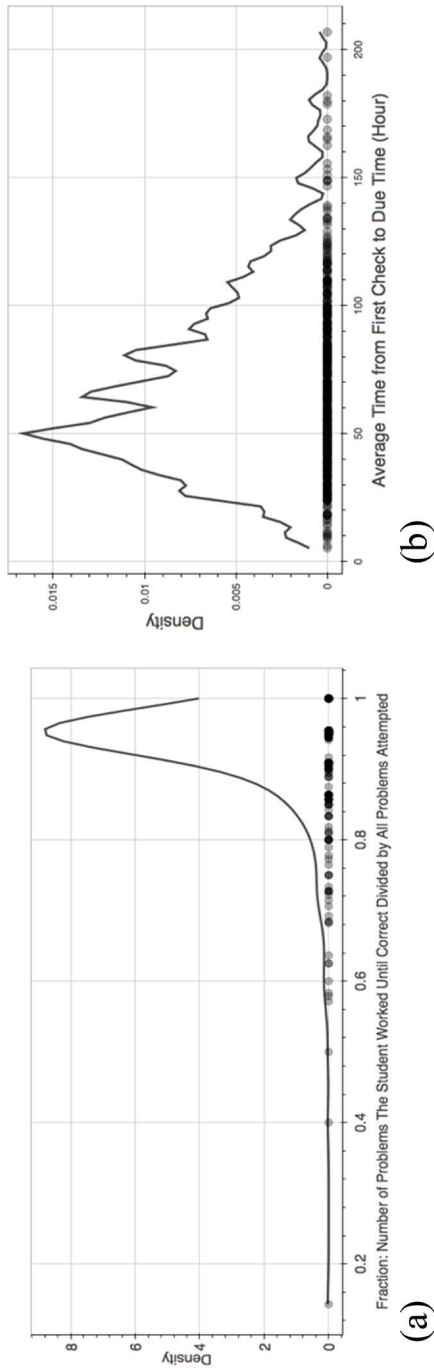


Fig. 2. a.) Density plot for the fraction of problems worked until correct divided by number of all problems attempted (variable 6 before weighting using difficulty levels); (b) Density plot for the average time across all homeworks from the first check to the homework due time for a given student (variable 8). [Each dot on the y = 0 axis represents a student at a specific x value.].

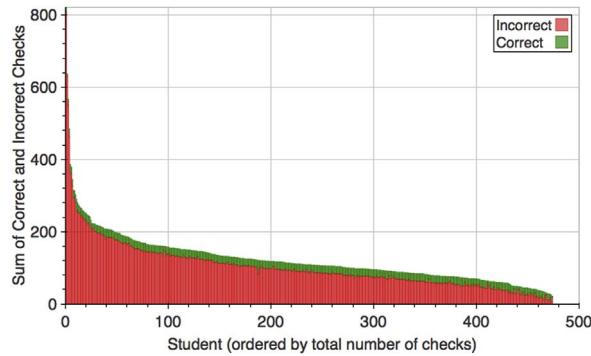


Fig. 3. Total number of correct and incorrect checks for each student over all 22 online homework problems.

Students took about 3 h between the first and second check for a given problem, on average, if they checked a second time. However, if we average across all of the checks instead of just the first two, it appears as though the first and second check are often farther apart than subsequent checks—the average time interval between all checks for the same problem is only about 4 min. This might indicate that students go through problems first in a more superficial manner to assess the assignment or try a first quick response. Although some students take much less time between checks (e.g., they might have quickly changed a response to reflect a possible algebraic mistake) and some take much more (e.g., they might have talked to other students), most do take enough time to use the feedback and reflect before resubmitting.

After generating the descriptive statistics, agglomerative hierarchical clustering allowed us to better understand the complex interactions among all the different ways the students engaged with the CAF as recorded by the online platform. We thus get a dendrogram structure of the variables as shown in Fig. 4. One characteristic of the ways students used the feedback that was immediately apparent was that the 30 behavioral variables were not independent from one another. The fact that variables such as these are not mutually exclusive is a common finding in research on learning both off and on line. Students' behaviors are mediated by various background factors such as prior knowledge, computer proficiency, metacognitive skills, and motivation. Therefore, two or more behavioral variables are likely mediated by a common set of hidden background factors and thus could be highly correlated with each other. Since we first extracted the variables in an effort to capture students' behaviors holistically without considering redundancy, it is not surprising that some variables are correlated with each other. For example, the number of lecture video accessed and the time spent viewing lecture videos are related, as is the students starting and then submitting their homework early.

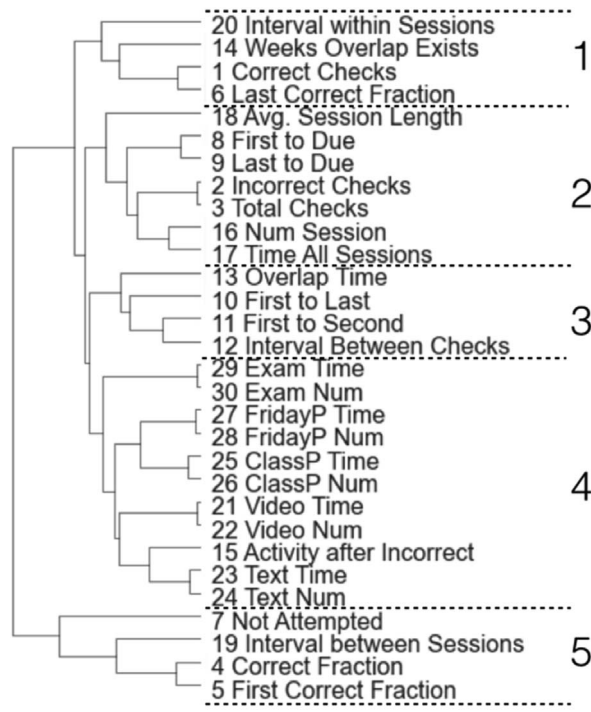


Fig. 4. Hierarchy of the 30 behavioral variables for online homework based on their Spearman correlation distances.

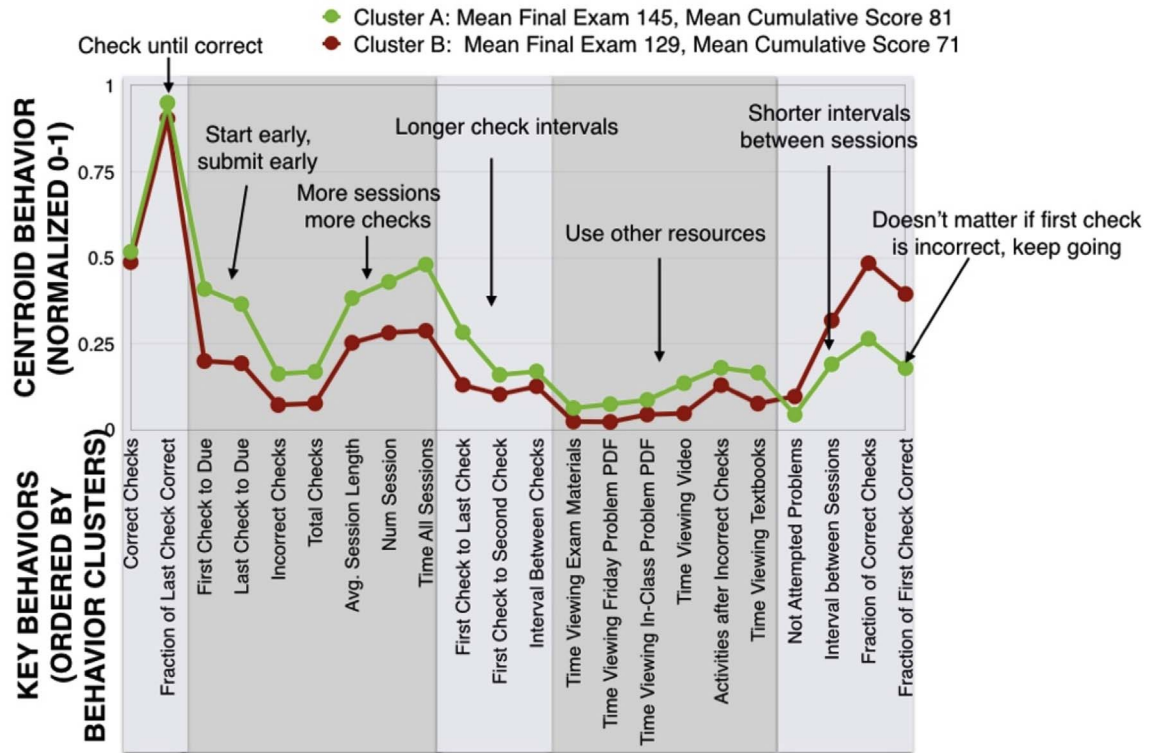


Fig. 5. Comparison of the centroids of the two student clusters for their behaviors on online homework problems (all variables normalized to 0–1).

We grouped the 30 behavioral variables into five basic clusters based on their Spearman rank correlation distances via agglomerative clustering (see Fig. 4). But as these clusters are inductively determined, they do not fit easily into separate concepts. However, broadly conceived, we might call them:

- (1) Persistence behaviors: measures such as the amount of time between problem-solving sessions, or the proportion of problems students worked on until correct.
- (2) Session checking behaviors: behaviors that characterize the way the students organized their study sessions, e.g., how long before the homework was due that we saw a check, or the number of distinct problem solving sessions for a given problem.
- (3) Inter-checking behaviors: more “micro” behaviors within a session that look at the checking behavior itself, e.g., the time between checks, and the time between the first and last check of a problem.
- (4) Reference resource checking behaviors: largely, the behaviors that described the ways in which students engaged with other resources (e.g., the videos, sample exams), including whether they engaged with these materials after getting an incorrect check.
- (5) Instrumental/basic engagement behaviors: measures that might describe very superficial engagement with the materials (e.g., not attempting problems or getting the problem right on the first check).

5.2. RQ 2(a): What patterns of engagement with simple corrective feedback comprise productive study behaviors and are associated with stronger performance?

Our next step in better understanding the way in which students’ behaviors with the CAF operated in a complex way was to identify behavioral clusters among the students. We used all 30 behavioral variables again to fully capture all the nuances in the patterns of engagement that different students displayed. Interestingly, despite being a diverse class with eight sections and over 400 students, we could identify just two distinctive and consistent clusters of students based on how they interacted with the CAF and other online resources. Cluster A contains 256 students while cluster B contains 218 students.

To more easily visualize the “average” behavior of these two groups of students, Fig. 5 compares the balancing points, or centroids, of the two clusters of students. Centroid behavior is not the behavior of one specific student; rather, we can trace the central tendency of the behavior of all students in that cluster. The variables in Fig. 5 are ordered and shaded based on the categorization of variables into the five groups we identified above. Importantly, as Fig. 5 also indicates, students in cluster A perform significantly better compared with students in cluster B in terms of both cumulative grades and final exam grades (both $p < 0.01$).

The data tell us about the family of behaviors that reflect self-regulated learning or productive struggle. Fig. 5 shows that these students engage in effective strategies such as checking multiple times, checking an answer until it is correct, and starting the

homework early and submitting it early. They also have longer, more frequent, and concentrated study sessions and use the various online resources more often.

With the knowledge that this overall pattern of behavior was related to higher performance, we could then proceed to our full multiple regression models. Our regression models allow us to parse out the variance attributable to specific student behaviors, while controlling for all other behaviors as well do students' demographics and educational background. We find that students' behavioral variables with the online materials can explain 25% of the variance in cumulative grades and 16% of the variance in the final exam grades. Appendix A gives regression results using the variance inflated factors (VIF)-selected variables as regressors for students' cumulative grades (columns 1 and 2) and final exam grade (columns 3 and 4). The VIF for selected variables range from 1.10 to 1.76, which means the influence on the standard error by the correlation among variables is at reasonable levels.

The time between the first check and the homework due time (variable 8) and *the time between the last check and the homework due time* (variable 9) are most strongly correlated to students' course performance. These students submitted the homework early compared with those who started and submitted it at the last minute. This may be due to their higher level of self-regulation, including more mature time organization strategies. Or it may be due to the increased amount of time they allowed for feedback, reflection, and integration. We see that the size of the correlation between the first check-to-due time and performance is greater for the student's cumulative grade than for the final exam score. In the regression analysis, more variables were used to predict the outcome variable of cumulative grade as opposed to final exam score. This is because the cumulative grade, as an outcome metric, encompasses numerous components of a student's effort and engagement with the class, rather than being a proxy measure of conceptual understanding like the final exam score.

Other variables that describe ways in which strong performers engage and potentially use feedback from the platform include the number of incorrect checks and the number of total checks for each student, which are highly positively correlated. In part, this is obvious because the total number of checks is the number of incorrect checks plus a near-constant number of correct checks.

We had previously thought that the outlier students who checked frequently might be demonstrating frequent guessing behaviors without carefully diagnosing their errors, and their performance might have been hurt by these behaviors. However, as we will show in the next section, students who had a high number of checks and a relatively low fraction of correct checks were not those who were checking carelessly; rather, they were those who spent more time engaging with the course platform, had longer time intervals between checks, worked until they got the correct answers, and, ultimately, performed well. We hypothesize that this behavior might indicate that even if a check is incorrect, the student receives feedback on their response and uses that feedback to learn. The positive relationship between incorrect checks and achievement may also be a proxy for the amount of effort the student puts in, which would contribute to performance.

We also find if students have longer, more frequent, and more concentrated study sessions, which is indicated by variables 17, 18, and 19 (*number of sessions, total session time, average session length, interval between sessions*), they tend to perform well in the course. These measures are also proxies for students' effort and interaction with the feedback. Finally, *the number of correct checks* (variable 1) and *the fraction of problems where students checked until correct* (variable 6) are positively correlated with student performance. Although students were not forced to answer until correct, those who did so have better outcomes than those who did not.

Variables in the third and fourth variable clusters, however, do not have a significant relationship with student performance. As we might expect, the third cluster (i.e., inter-checking behavior) fluctuates with the second cluster (i.e., session checking behaviors), but is not related to performance on its own. The fact that the fourth cluster (i.e., resource checking behaviors) does not correlate with student performance may indicate that the use of other online resources in the course platform is not a good measure for engagement. We believe this may be because in this blended learning context, students have many other resources to help them such as peer discussion, tutoring sessions, and in-class problem solving sessions. This reflects one of the major differences between blended learning and entirely online courses.

In sum, although the online homework grades constitute only 2% of the cumulative grade in the course, we see that engaging with the online homework and immediate feedback is hugely beneficial to students' success in the course. These results are further strengthened by our multiple regression analysis: We find that some key CAF behaviors are significant predictors of student performance, even when controlling for important correlates like first-generation college status or incoming SAT II science scores.

5.3. RQ 2(b): What patterns of engagement with simple corrective feedback negatively correlated with performance, including copying?

There are several behaviors that we can see students engaging in that do not lead to higher performance. For example, *the interval between sessions* (variable 19) is negatively related to students' performance. This may be because if students wait too long to resume their study sessions, they might have forgotten the feedback they obtained from previous sessions. Therefore, they lose the opportunity to benefit from reflecting on what they have already understood and using it to further their learning. Variable 7 (*number of not attempted online homework problems*) is negatively correlated with course performance in our bivariate correlation results, which helps support findings in educational research that time on task is important.

We had thought that students getting *the correct answer on their first try* (variable 5) might be a proxy for stronger prior knowledge of the course content. Based on previous work, this would likely > correlate highly positively with subsequent course performance. Surprisingly, our correlation analysis shows the opposite. Fig. 6 shows that those who got the correct answers on the first try tended to be those who started the homework late and submitted the homework in the last two days preceding the due date. Although these students got most of their online homework problems correct (variable 4) and thus had high online homework scores, their final exam and cumulative grades appear to have been hurt by this behavior. Perhaps these students are copying answers from other students at the last minute rather than using knowledge of the subject. At minimum, these students had very little engagement with the material

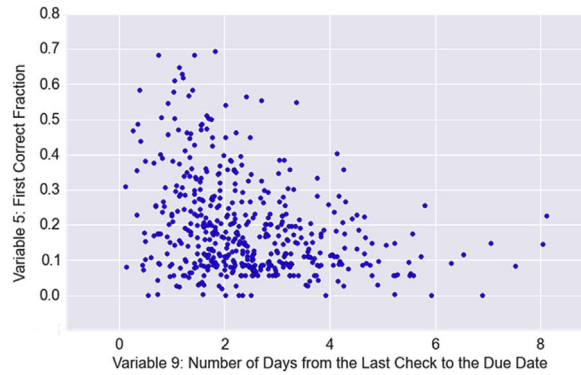


Fig. 6. Students who submit the homework late have higher first correct fraction.

and the immediate feedback.

Although infrequent, we did find evidence of behaviors that might indicate copying. These behaviors were negatively correlated to success in the course. However, we did not find much evidence for other “gaming” behaviors. One reason might be that the students were not forced to answer the homework questions until correct, so there was no need to game the system. They could simply give up and disengage.

6. Discussion

This study utilizes rich quantitative data to understand how students interact with immediate feedback on their online homework assignments in a blended learning course. The feedback is simple and corrective: once the students submit an answer to a problem, the checkable answer feature (CAF) responds with either a green tick mark or a red X to tell them if they are right or wrong. Although a substantial body of research exists on the importance of feedback to the learning process, clickstream data generated by platforms like edX allow researchers to see more closely how students make use of feedback. We can observe how interacting with the CAF helps student master concepts in mechanics, and how that behavior correlates with better scores on the final exam and higher grades in the course. In the broadest sense, this research reinforces what educational research has shown consistently: “effort” and “engagement” correlate robustly with performance.

We identify two distinctive behavioral clusters among students. One group, Cluster A, displayed more productive and persistent behaviors than Cluster B, and, subsequently, those students earned higher grades. They engaged with the CAF in similar ways: they were more organized in approaching the homework and they spent more time on it. More specifically, they started their homework early and submitted it early, checked answers multiple times, continued until the platform confirmed they got the correct answer, and worked on the assignment regularly with shorter intervals between problem solving sessions. Most students in the course attempted at least 20 out of 22 problems, but the ways in which Cluster A students approached and carried out assignments significantly correlated with their success in the course.

The behavioral variable that relates most strongly to achievement is beginning homework early, as indicated by the first check. Cluster A students did just that, and we hypothesize they worked more studiously and were able to receive and incorporate feedback more effectively. We also saw that these students were tenacious: they kept working on a problem until they got the right answer, even if their first attempts were wrong. In fact, more checks correlate with higher performance, so even if these students never got the right answer, their persistence paid off. Students who started and submitted their assignment relatively later and seemed to do the homework in a more hurried or superficial way (e.g., they got the answer right on their first attempt) did not have as much time/engagement with the course materials and the CAF. As we show, they did not achieve grades as high as their classmates who worked more conscientiously.

These results have several implications. First, this study is an example of how data generated at a fine grain by an online platform can help corroborate educational research based on more traditional data collection and analysis methodologies. We show how questions generated by previous research can be combined with inferential statistics and newer methods of analysis from data mining, as we alternated between deductive and inductive approaches. For example, our results reinforce numerous findings in the literature on the role of feedback on learning, most generally that immediate feedback can help enormously in mastering new domains of knowledge.

Similarly, the advantageous behaviors we see demonstrated by Cluster A students help substantiate what prior research has found on the importance of self-regulated learning and productive struggle. Students in Cluster A exhibit behaviors that seem to indicate they are taking control of their learning and using the CAF in ways that furthered it. For example, even though students were not required to work until they got an answer right, those who persisted performed better. In fact, contrary to what the instructional team and the educational researchers had hypothesized, we found that students who got a higher proportion of incorrect first answers, but kept working, did better overall.

Our study also provides insight into how instructors can guide students to success in the course. Faculty should reward students for “trying,” encouraging them to work on a problem until they understand it. Instructors should stress that often effort is more

beneficial for achievement than efficiency. Our results support this point because we know that continuing to check (derived, for example, from the number of incorrect checks) and working on the problem (derived from the time between first check and last check, as well as overall session time) correlated more highly with achievement than getting the correct answer with the fewest possible number of tries. Teachers should urge students to organize more frequent and concentrated study sessions since, as we show, shorter intervals between sessions resulted in better grades. We realize these are not entirely new findings in educational research. But we hope that the data from the large data set we had available to us, combined with innovative methods of analysis, can add to an understanding of how best to advise students on study strategies. In the same vein, we hope to provide more ammunition for teachers (and, for that matter, parents) who want to emphasize the importance of being tenacious, working through hard-to-learn material, and solving problems that at first can seem intractable. As we report in our findings, although online homework problems only constituted 2% of the cumulative grade in the course, student behaviors with online homework problems explain 25% of the variance in the cumulative grades.

We do recognize that our research is limited in that we do not know if the students we see in Cluster A began the class with strong skills in self-regulation or are conscientious and persistent by nature. Therefore, we cannot say with certainty how much the existence of the CAF contributed to the behaviors we observed. We did not have direct measures of or strong proxies for how much prior knowledge students had as they began the course, or what their experience with course content had been previously. And, of course, we could not actually observe everything that students were doing in between checks either online, on a different website, or off the platform. This limits our ability to completely affirm or explain our hypotheses about why we observe the relationships we find in our study. These would be fruitful areas for future research.

7. Conclusion

Online platforms can now provide researchers with data that indicate what resources students are using, how they come to understand concepts, and what behaviors reflect effective study strategies. Thus, these platforms allow us to study learning at a deeper level than we have been able to. In this study, we examine how students use very simple online task-level feedback to master material that they may not have encountered before, recognizing that feedback is integral to the learning process. The data show us that students who take advantage of that feedback and use techniques associated with self-regulation and productive struggle have stronger results in the course. These findings can be used by instructors to design better features of online learning platforms, integrate online activities with the face-to-face components of their courses, and guide students on using online material advantageously.

Acknowledgements

We would like to recognize the contributions of undergraduate researcher Xingyu Zhou (Purdue University) in data cleaning and understanding the data generating mechanism of the platform. We would like to thank the instructors and researchers at the university studied for their participation and help with the ongoing investigation.

We would like to acknowledge the support and feedback from the first author's doctoral committee members Dr. Ruth Streveler and Dr. Sean Brophy.

This research is based upon work supported in part on funding provided by the university and in part by the National Science Foundation (NSF) under #1454558. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of NSF.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.compedu.2017.09.013>.

References

- Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? *Intelligent tutoring systems* (pp. 292–303). Springer. Retrieved from http://link.springer.com/chapter/10.1007/3-540-45108-0_33.
- Angus, S. D., & Watson, J. (2009). Does regular online testing enhance student learning in the numerical sciences? Robust evidence from a large data set. *British Journal of Educational Technology*, 40(2), 255–272.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004). Off-task behavior in the cognitive tutor classroom: When students game the system. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 383–390). ACM.
- Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, 106(3), 410.
- Beichner, R. J., & Saul, J. M. (2003). Introduction to the SCALE-UP (student-centered activities for large enrollment undergraduate programs) project. *Proceedings of the international school of physics "enrico fermi."* Varenna, Italy.
- Beichner, R. J., Saul, J. M., Abbott, D. S., Morse, J. J., Deardorff, D., Allain, R. J., ... Risley, J. S. (2007). The student-centered activities for large enrollment undergraduate programs (SCALE-UP) project. *Research-based Reform of University Physics*, 1(1), 2–39.
- Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, 79(3), 1243–1289.
- Bernard, R. M., Borokhovski, E., Schmid, R. F., Tamim, R. M., & Abrami, P. C. (2014). A meta-analysis of blended learning and technology use in higher education: from the general to the applied. *Journal of Computing in Higher Education*, 26(1), 87–122.
- Boekaerts, M. (1999). Self-regulated learning: Where we are today. *International Journal of Educational Research*, 31(6), 445–457.
- Bransford, J., Brown, A. L., & Cocking, R. R. (1999). *How people learn: Brain, mind, experience, and school*. Washington, D.C.: National Academies Press.
- Breslow, L. (2010). Wrestling with pedagogical change: the TEAL initiative at MIT. *Change: The Magazine of Higher Learning*, 42(5), 23–29.
- Chung, G. K., Shel, T., & Kaiser, W. J. (2006). An exploratory study of a novel online formative assessment and instructional tool to promote students' circuit problem

- solving. *The Journal of Technology, Learning and Assessment*, 5(6), Retrieved from <http://ejournals.bc.edu/ojs/index.php/jtla/article/view/1645>.
- Clariana, R. (1990). A comparison of answer until correct feedback and knowledge of correct response feedback under two conditions of contextualization. *Journal of Computer-based Instruction*, 17(4), 125–129.
- Crisp, V., & Ward, C. (2008). The development of a formative scenario-based computer assisted assessment tool in psychology for teachers: The PePCCA project. *Computers & Education*, 50(4), 1509–1526.
- DeBoer, J., Ho, A., Stump, G. S., Pritchard, D. E., Seaton, D., & Breslow, L. (2013). Bringing student backgrounds online: MOOC user demographics, site usage, and online learning. *Engineer*, 2, 0–81.
- Dihoff, R. E., Brosvic, G. M., & Epstein, M. L. (2003). The role of feedback during academic testing: The delay retention effect revisited. *The Psychological Record*, 53(4).
- Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2004). Provision of feedback during preparation for academic testing: Learning is enhanced by immediate but not delayed feedback. *Psychological Record*, 54(2), 207–232.
- Earley, P. C., Northcraft, G. B., Lee, C., & Lituchy, T. R. (1990). Impact of process and outcome feedback on the relation of goal setting to task performance. *Academy of Management Journal*, 33(1), 87–105.
- Epstein, M. L., Lazarus, A., Calvano, T., Matthews, K., Hendel, R., Epstein, B., et al. (2010). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, 52(2), 187–201.
- Feldman, A., & Capobianco, B. M. (2008). Teacher learning of technology enhanced formative assessment. *Journal of Science Education and Technology*, 17(1), 82–99.
- Gijbels, D., Dochy, F., Van den Bossche, P., & Segers, M. (2005). Effects of problem-based learning: A meta-analysis from the angle of assessment. *Review of Educational Research*, 75(1), 27–61.
- Google (2015). Google analytics cookie usage on websites. Retrieved October 8, 2015, from <https://developers.google.com/analytics/devguides/collection/analyticsjs/cookie-usage>.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>.
- He, W., Gajski, D., Farkas, G., & Warschauer, M. (2015). Implementing flexible hybrid instruction in an electrical engineering course: The best of three worlds? *Computers & Education*, 81, 59–68. <https://doi.org/10.1016/j.compedu.2014.09.005>.
- Interactive Advertising Bureau (2009). IAB - guidelines, specifications & best practices. Retrieved October 8, 2015, from <http://www.iab.net/guidelines>.
- Koh, L. C. (2008). Refocusing formative feedback to enhance learning in pre-registration nurse education. *Nurse Education in Practice*, 8(4), 223–230.
- Kulhavy, R. W., & Anderson, R. C. (1972). Delay-retention effect with multiple-choice tests. *Journal of Educational Psychology*, 63(5), 505.
- Kulik, J. A., & Kulik, C.-L. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research*, 58(1), 79–97.
- Lack, K. A. (2013). *Current status of research on online learning in postsecondary education*. Ithaka. Retrieved from <http://sr.ithaka.org/research-publications/current-status-research-online-learning-postsecondary-education>.
- Lin, Q. (2008). Preservice teachers' learning experiences of constructing e-portfolios online. *The Internet and Higher Education*, 11(3), 194–200.
- Maier, U., Wolf, N., & Randler, C. (2016). Effects of a computer-assisted formative assessment intervention based on multiple-tier diagnostic items and different feedback types. *Computers & Education*, 95, 85–98.
- Mazingo, D. E. (2006). *Identifying the relationship between feedback provided in computer-assisted instructional modules, science self-efficacy, and academic achievement*. University of Colorado Health Sciences Center. Retrieved from <http://dl.acm.org/citation.cfm?id=1329856>.
- Morris, L. V., Finnegan, C., & Wu, S. (2005). Tracking student behavior, persistence, and achievement in online courses. *The Internet and Higher Education*, 8(3), 221–231. <https://doi.org/10.1016/j.iheduc.2005.06.009>.
- Nariciss, S., & Huth, K. (2006). Fostering achievement and motivation with bug-related tutoring feedback in a computer-based training for written subtraction. *Learning and Instruction*, 16(4), 310–322.
- Northcutt, C. G., Ho, A. D., & Chuang, I. L. (2016). Detecting and preventing “multiple-account” cheating in massive open online courses. *Computers & Education*, 100, 71–80. <https://doi.org/10.1016/j.compedu.2016.04.008>.
- Papert, S. (1993). *Mindstorms: Children, computers, and powerful ideas*. Da Capo Press.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, D.C.: National Academy Press, Board on Testing and Assessment, Center for Education, Division of Behavioral and Social Sciences and Education.
- Pennebaker, J. W., Gosling, S. D., & Ferrell, J. D. (2013). Daily online testing in large classes: Boosting college performance while reducing achievement gaps. *PLoS One*, 8(11), e79774. <https://doi.org/10.1371/journal.pone.0079774>.
- Persky, A. M., & Pollack, G. M. (2008). Using answer-until-correct examinations to provide immediate feedback to students in a pharmacokinetics course. *American Journal of Pharmaceutical Education*, 72(4).
- Pintrich, P. R., & de Groot, E. V. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82(1), 33–40. <https://doi.org/10.1037/0022-0663.82.1.33>.
- Rayyan, S., & Belcher, J. (2014, November). 8.02 TEAL + x: Students say “Yes” to MITx in 8.02 TEAL. *MIT Faculty Newsletter*, XXVII(2), Retrieved from http://web.mit.edu/fnl/volume/272/rayyan_belcher.html.
- Rokach, L. (2010). A survey of clustering algorithms. *Data Mining and Knowledge Discovery Handbook* (pp. 269–298). Springer.
- Ruiperez-Valiente, J. A., Alexandron, G., Chen, Z., & Pritchard, D. E. (2016). Using multiple accounts for harvesting solutions in MOOCs. *Proceedings of the third (2016) ACM conference on learning @ scale* (pp. 63–70). New York, NY, USA: ACM. <https://doi.org/10.1145/2876034.2876037>.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>.
- Silva, E., & White, T. (2013). *Pathways to improvement: Using psychological strategies to help college students master developmental math*. Carnegie Foundation for the Advancement of Teaching.
- Sorensen, E. K., & Takle, E. S. (2005). Investigating knowledge building dialogues in networked communities of practice. A collaborative learning endeavor across cultures. *Interactive Educational Multimedia: IEM*, 10, 50–60.
- Van der Kleij, F. M., Eggen, T. J., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263–272.
- Van der Kleij, F. M., Feskens, R. C., & Eggen, T. J. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of Educational Research*, 85(4), 475–511.
- Van der Pol, J., Van den Berg, B. A. M., Admiraal, W. F., & Simons, P. R.-J. (2008). The nature, reception, and use of online peer feedback in higher education. *Computers & Education*, 51(4), 1804–1817.
- Vedel, A. (2014). The big five and tertiary academic performance: A systematic review and meta-analysis. *Personality and Individual Differences*, 71, 66–76.
- Vonderwell, S., Liang, X., & Alderman, K. (2007). Asynchronous discussions and assessment in online learning. *Journal of Research on Technology in Education*, 39(3), 309–328.
- Vygotsky, L. S. (1986). In A. Kozulin (Ed.). *Thought and language*. The MIT Press Rev'd.
- Wang, T.-H., Wang, K.-H., & Huang, S.-C. (2008). Designing a web-based assessment environment for improving pre-service teacher assessment literacy. *Computers & Education*, 51(1), 448–462.
- Wei, H., Peng, H., & Chou, C. (2015). Can more interactivity improve learning achievement in an online course? Effects of college students' perception and actual use of a course-management system on their learning achievement. *Computers & Education*, 83, 10–21. <https://doi.org/10.1016/j.compedu.2014.12.013>.
- Wilson, J. M. (1994). The CUPLE physics studio. *The Physics Teacher*, 32(9), 518–523.
- Wolsey, T. D. (2008). Efficacy of instructor feedback on written work in an online program. *International Journal on ELearning*, 7(2), 311.
- Zhao, Y., & Breslow, L. (2013). Literature review on hybrid/blended learning. Retrieved from http://tll.mit.edu/sites/default/files/library/Blended_Learning_Lit_Review.pdf.
- Zheng, B., & Warschauer, M. (2015). Participation, interaction, and academic achievement in an online discussion environment. *Computers & Education*, 84, 78–89. <https://doi.org/10.1016/j.compedu.2015.01.008>.
- Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist*, 25(1), 3–17.