Analyzing the relative learning benefits of completing required activities and optional readings in online courses

Paulo F. Carvalho
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
pcarvalh@cs.cmu.edu

Min Gao
Beijing Normal University
19 Xin-jie-kou Wai St,
Beijing 100875, P. R. China
bnugm2014@163.com

Benjamin A. Motz
Indiana University
1101 E. 10th St.
Bloomington, IN 47405
bmotz@indiana.edu

Kenneth R. Koedinger
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA 15213
koedinger@cmu.edu

ABSTRACT

Students who actively engage with learning materials, for example by completing more practice activities, show better learning outcomes. A straightforward step to stimulate this desirable behavior is to require students to complete activities and downplay the role of reading materials. However, this approach might have undesirable consequences, such as inflating the number of activities completed in a short period of time until maximum performance is achieved ("gaming the system"). In this paper, we analyze the relative benefits of completing activities vs. readings for learning outcomes in an online course that required students to perform practice activities. The results show that students who read more pages have better learning outcomes than students who completed more activities. This pattern of results holds even when considering different measures of active engagement but is reversed when considering only activities classified as effective active engagement by a "gaming behavior" classifier. Overall, these results suggest that, when completing activities is required, students benefit from complementing the activities with optional readings. One possibility is that completing optional readings can be an active learning activity in itself, driven by students who are going beyond the minimum requirement, and actively seeking further information and robust feedback that complements the activities.

Keywords

Active learning; online learning; "game the system" classifiers

1. INTRODUCTION

Students learn better when they engage in active learning [11,19]. Yet, much instructional practice emphasizes passive learning such as reading text, attending lectures, and watching videos. Contrary to evidence of the clear benefits of active learning, students (and a surprisingly high number of instructors) feel that passive strategies such as re-reading are useful study methods [16]. This disconnect between evidence and practice highlights the need to develop active learning practices that are grounded in empirical evidence and can support effective learning. In this paper we investigate the positive benefits of active learning in an online course and the effect of encouraging students to engage in pre-determined active learning activities.

Online courses might, by their nature, lead to fewer active learning practices. For example, online courses often rely on text and videos to convey information, typical passive learning practices. However, although video- or text-based online courses are common, previous work by Koedinger and collaborators has suggested that greater engagement with practice activities in online courses is a better predictor of improved learning than greater engagement with video or text materials [7,13,14]. In light of this research, one suggestion would be that more activities should be included in online courses, and students should be encouraged to complete them. However, two problems arise from trying to implement this suggestion: how to encourage students to complete activities and what type of activities to use.

Effective self-regulation skills play an important role for successful learning in in-person instruction [5], as well as in online courses [4,12]. With the added autonomy afforded by online courses compared to in-person instruction, students who lack appropriate self-regulation skills or try to complete the course with the minimum amount of time and effort might not perform as well. Thus, it is important to encourage students who might not otherwise engage in active learning to do so [5], both because it might be more time consuming and effortful than passive learning techniques but also because engaging in active learning stimulates self-regulation and accurate learning calibration [10]. One straightforward way to do so in online courses is to include multiple practice activities in each online lesson and make performance in the activities count towards the students' grade. This suggestion is not without its challenges, however. While this approach might encourage students who otherwise would not complete the activities to do so, it might be problematic if regulating one's own activities is a critical ingredient in the learning process. Indeed, previous research on other cognitive approaches to improve learning, have repeatedly shown a difference in outcomes between when students are in control of their study and when they are not [6,8]. Another issue is related to "gaming the system" behaviors. Making activity completion explicitly related to grade outcomes, might lead students to attempt to exploit the activities not as learning devices, but a way to quickly achieve better grades [1,2].

There is also the issue of how the activities should be designed. Previous research investigating the positive effect of completing more activities in online courses looked at courses using activities that not only were optional, but also included extensive feedback, both for correct and incorrect responses. It is possible that the characteristics of the activity used play a role on whether they contribute to improved learning [15].

With these questions in mind, in the current study we investigated the relative benefits of completing activities and reading textbook materials for learning outcomes in an online course. The main research questions were (1) whether completing more practice activities would contribute to better learning outcomes than accessing more textbook pages when students are required to complete the activities and are provided minimal feedback in the activities, and (2) whether we could detect students' active engagement in the activities and distinguish it from 'gaming the system' behaviors such as completing the same activity multiple times quickly until a high score was achieved.

We use data from an exclusively online course taught at Indiana University. This course had a few characteristics that made it particularly relevant for the current research questions: (1) it included many practice activities in each unit, (2) the activities were required, graded and made up a large part of the students' final grade, but (3) students were allowed to complete the activities as many times as desired, (4) the activities included only correctness feedback, and (5) the textbook materials were separated from the rest of the course materials.

We start by analyzing the relative benefits of completing more activities vs. accessing more textbook pages as in previous research. Next, we investigate possible markers of "active learning" engagement that might influence the relative benefit of completing more activities on learning outcomes that help identify behaviors to use in the classifiers. Finally, we developed two classifiers to detect, among all activity completion attempts, which ones might involve "active learning" behaviors, and which might involve "gaming the system" behaviors. We then use measures derived from these classifiers to evaluate the relative benefits of more active completions of activities vs. accessing more textbook pages.

2. DATA AND METHODS

We used data from two semesters of an online introductory psychology course at Indiana University (N = 247 and N = 492, respectively). All students enrolled in the course were undergraduate students at one of the campuses of Indiana University taking the course for credit. All students' rights as research participants were protected under a protocol approved by the local review board and were informed in the course syllabus that their data would be analyzed.

2.1 Course Description

Table 1. Number of assigned activities and textbook readings available in Units 2-7.

Unit	Number of lesson activities	Number of Textbook readings
2 – Methods	34	68
3 – Neuroscience	24	46
4 – Perception	30	43
5 – Memory	19	37
6 – Learning	20	44
7 - Cognition	18	36

The course was developed by the third author and delivered through Canvas. The course had seven units, but the first unit was purely introductory (there was no quiz at the end of the first unit) and is not included in the present analysis, leaving six content units for the current study (listed in Table 1). All units started with a short video from the instructor presenting an overview of the main topics of the unit. Moreover, every unit contained a different number of lessons, and within each lesson a different number of pages, each dedicated to a sub-topic. Every page contained an abbreviated summary of the main points of the sub-topic, links to the relevant readings of the online textbook, and lesson activities. Some pages also included videos and demonstrations. The number of lesson activities and textbook pages varied from unit to unit (see Table 1).

2.1.1 Lesson activities

Students were required to complete all the practice activities within the lessons of all units, using a custom LTI-based assessment platform installed in Canvas (Quick Check; https://github.com/IUeDS/quickcheck). Performance on these activities accounted for 45% of the students' final grade. Lessons were scheduled, and activities had to be completed within the scheduled time for each lesson. Students were allowed to complete the activities as many times as desired before or after the lesson completion deadline. Only their highest score before the lesson completion deadline was considered for their grade, and activities completed after the deadline never counted toward the students' grade. The lowest four aggregate lesson scores were automatically dropped. Aggregate lesson scores indicate the scores of all activities in the same lesson.

Lesson activities covered the content of the specific lesson they were assigned to and varied in format across lessons, including, for example, multiple-choice and graph interpretation activities. An example of two activities is included in Figure 1.

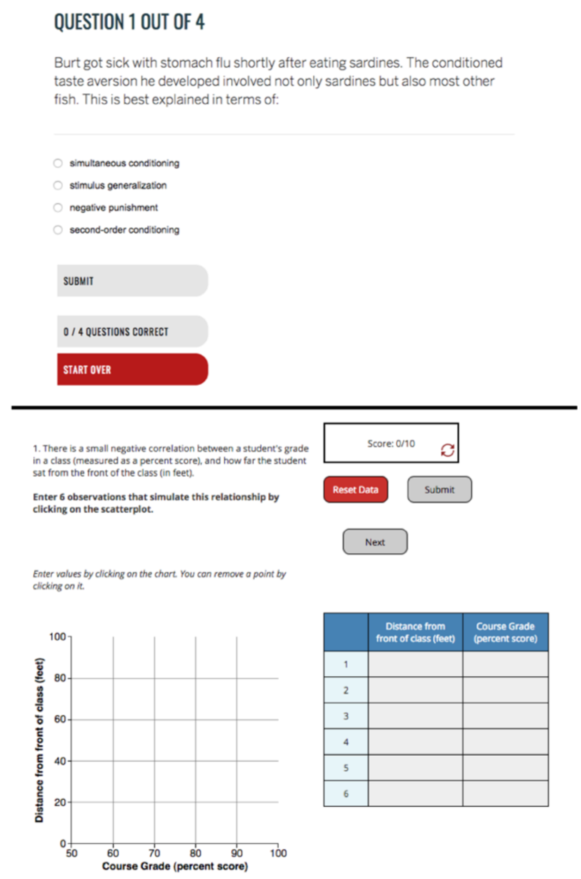

Figure 1. Examples of two lesson activities.

2.1.2 *Textbook readings*

The course used an online version of a commercially available Introductory Psychology textbook through the Unizin platform (eText). All students had access to the eText as part of their enrollment in the course. The relevant pages of the textbook for each topic covered in each page of every lesson across all units was provided in the course (see Figure 2 for an image). Students could access the eText at any point, including during the exams. Importantly, reading the eText was not incentivized or rewarded with points.

Figure 2. Example of link to eText pages on lesson (top right) and corresponding eText page in new window (bottom).

2.1.3 *Quizzes*

At the end of every unit, students completed a timed quiz online. Students could only attempt the quiz once within the time-frame allotted. Quizzes included a series of multiple-choice questions randomly chosen for each student from a larger pool. Quizzes accounted for 40% of the student's final grade and the lowest quiz grade was automatically dropped.

2.1.4 *Reflection activities*

Finally, students also completed a reflection activity for each unit. These activities were a writing assignment designed to help students think about the course materials for that unit in more depth. These assignments were due when the quiz for the unit became available and accounted for 15% of the students' final grade. The lowest score was automatically dropped.

2.2 Data

Detailed logged information was collected for this course. We analyzed information regarding when each lesson activity was attempted and how many times, how many eText pages were accessed and when, as well as scores on the lesson activities and the quizzes. The logged information allowed us to determine how long students took completing activities, but not how long they spent reading.

2.3 Model building

In order to compare the relative effect of different student behavior we normalized all measures by converting them to z-scores. Unless otherwise stated, we used mixed effects regression models to investigate the effect of different student behaviors on quiz scores. The baseline model included number of activities completed and the number of eText pages accessed. We predict quiz performance for each quiz, considering only behaviors that took place *before* the quiz was made available to the students:

$$\text{ZquizScores} \sim \text{Zactivities} + \text{Zpages} + (1|\text{student}) + (1|\text{quiz}) \quad (1)$$

This base model includes activities completed before the corresponding due date or after the due date as long as it was before the start of the quiz period. Considering only activities completed before the corresponding due date does not change any of the result patterns reported here. To help establish potential causal relations, we also ran the same baseline model predicting quiz grades using only behaviors that took place *after* the quiz was made available. We included student and quiz number as random effects in all models. We extracted different student behavior features and added them to the baseline model to infer the relative benefit of doing and reading using different properties of doing (e.g., time and accuracy). We use *chi-square* to compare models.

In addition, we developed two different classifiers to identify active engagement with the activities and discriminate it from possible "gaming the system" behaviors by the students (see below for details). We then include the 'gamer' classifier as an added predictor to the baseline model.

3. RESULTS

We started by running all analyses separately for each semester. All patterns were similar across both datasets; thus, we combined the two datasets into a single dataset for all analyses reported below. For brevity, we focus only on quiz performance as outcome measure, a similar pattern of results was found when considering the reflection activities as outcome measure.

3.1 Description of main variables

3.1.1 *Lesson activities*

Students completed an average of 74 activities before the quiz (Median = 71, SD = 40), and took on average 201 minutes (Median = 114, SD = 246) doing so. Only an average of 22% of these activities were completed after the activity due date but before the quiz, therefore in all subsequent analyses we consider any activity completed before the start of the quiz, regardless of the activity specific due date. After the corresponding quiz start date, students completed an average of 17 activities (Median = 5, SD = 25), taking on average 20 minutes doing so (Median = 0, SD = 61).

3.1.2 *eText*

Students opened an average of 22.5 eText pages before the corresponding quiz (Median = 5, SD = 35) and 12 pages after the corresponding quiz was made available (Median = 5, SD = 17).

3.1.3 *Quizzes*

The mean quiz score was 23.88 (Median = 25, SD = 4.80) out of 30 possible points. The distribution of quiz scores as a function of number of activities completed and number of pages opened before the quiz is presented in Figure 4.

3.2 Base models: Relative benefit of doing and reading before the quiz start date

3.2.1 *Behavior before the quiz start date*

Accessing more eText pages before the quiz being made available predicts better quiz performance, $\beta = 0.14$, t (3689) = 8.07, $p <$ .0001. Conversely, completing more activities before the quiz was made available predicts worse quiz performance, $\beta = -0.04$, t (3733) = -2.06, $p = .039$.

Overall, we do not see a "doer effect", i.e., that completing more practice activities improves learning outcomes to a larger degree than completing more readings.

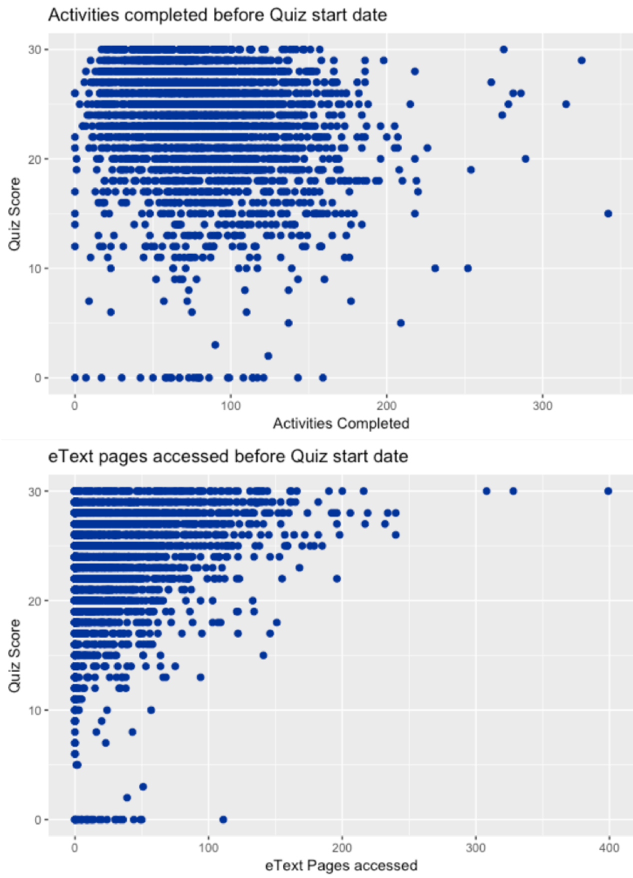

Figure 3. Distribution of quiz scores as a function of number of lesson activities (top panel) and eText pages (bottom panel) accessed before the quiz start date.

3.2.2 *Behavior after quiz start date*

Accessing more eText pages after the quiz was made available predicts better quiz performance, $\beta = 0.05$, $t\ (3590) = 2.94$, $p <$ .0001, potentially because students were using the eText to complete the quiz. Completing more activities after the quiz was made available also predicts better quiz performance, $\beta = 0.18$, t (3810) = 12.38, $p < .0001$.

Thus, completing more activities after the quiz was made available had a larger effect on outcomes than accessing more eText pages, contrary to what we saw when analyzing behaviors before the quiz was made available.

3.3 Time and performance models

The learning benefit of completing more activities is likely to be connected with active engagement with the activities. However, the requirement to complete activities and the fact that performance on these activities directly affected students' grades might have led students to complete the activities multiple times in quick succession for maximum performance (a "gaming the system" behavior). This is potentially a different type of activity engagement that would not lead to a doer effect. To test this hypothesis, we created models that include measures potentially more related to active engagement: (a) time working on activities, (b) average performance across all activity attempts, and (c) best performance weighted by number of activity attempts. We compare models including each of these measures as added predictors with the baseline model for behaviors before the quiz described above.

3.3.1 *Time working on activities*

Spending more time working on the activities before the quiz has a positive impact on quiz performance, $\beta = 0.09$, $t\ (3818) = 5.47$, $p <$ .0001. Moreover, compared to the baseline model, the activity time model provided a significantly better fit to the data, $\chi^2 = 29.34$, $p <$ .0001 (see Table 2).

3.3.2 *Average performance on activities*

Higher average performance on the activities completed before the quiz is also related to higher quiz performance, $\beta = 0.04$, $t\ (3796) =$ 2.604, p = .009. Compared to the baseline model, the activity performance model provided a significantly better fit to the data, $\chi^2 = 6.64$, $p = .01$ (see Table 2).

3.3.3 *Number-weighted best performance*

Only the highest score across all attempts was considered for student final grade. Therefore, it is likely that students who achieved higher scores with less attempts were more actively engaged in the activities than students who achieved higher scores with more attempts. The latter group was likely to be attempting to achieve a high score by completing the activity multiple times without attending to the actual question or feedback. Achieving highest scores in less attempts predicted better quiz results, $\beta =$ 0.04, t (3365) = 3.10, p = .002. This model also provides a significantly better fit to the data compared to the baseline model, $\chi^2 = 9.46$, $p < .002$ (see Table 2).

3.4 Detecting effective activity use

The findings of the previous section suggest that not all activity completion is active learning, and some might reflect "gaming the system" behaviors. This raises the important question of being able to distinguish effective active learning in activity use from other uses. From the previous analyses, we concluded spending more time, being more accurate across all attempts and achieving highest score with less attempts all predict better quiz performance and provide better fit to the data. Using these findings, we created two classifiers of "gaming the system" behaviors. One that takes only attempt duration into account and another that takes into account not only duration but also accuracy of each attempt.

Table 2. Summary of regression models used to evaluate the benefits of doing and reading in the online course.

Model	Number activities	eText pages	Added predictor	AIC	BIC
Baseline (before quiz start)	-0.04*	0.14***	-	9404.7	9442.2
Baseline (after quiz start)	0.18***	0.05**	-	9311.3	9348.8
Time working on activities	-0.05**	0.12***	0.10***	9377.4	9421.1
Average performance on activities	-0.02	0.14***	0.04**	9400.1	9443.8
Number-weighted best performance on activities	-0.03	0.14***	0.04**	9397.3	9441.0
Effective active learning activity use (duration-based)	0.29	0.14***	-0.33	9404.0	9447.8
Effective active learning activity use (duration+accuracy)	-0.21**	0.14***	0.17*	9400.1	9443.8

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

3.4.1 Duration-based classifier

Using the raw attempt log for each activity for each student, we determined whether each attempt was faster than what is "normal" for that student by considering that students' median time completing similar activities. An attempt was considered as not effective active engagement if it was shorter than the median of all attempts for that student for the same activity. Thus, in essence, this classifier positions each attempt as too quick to be likely to involve active engagement, based on how long students often take to complete similar activities, and is consistent with our findings in the previous section that time is a better predictor of effective activity use. This classifier identified approximately 47% of attempts before the quiz as not involving active engagement.

3.4.2 Duration+accuracy classifier

Using the raw attempt log for each activity for each student, we determined whether each attempt was faster and less accurate than what is "normal" for that student by considering that students' median time completing all activities and their median accuracy. The previous analyses suggested that accuracy in the activities was also a good predictor of effective activity use. Thus, the premise for this classifier was that if students were merely completing the same activity multiple times by randomly varying their answers until reaching high scores, one would expect that it would involve multiple short attempts with low accuracy. Active engagement, on the other hand, would involve longer attempts with higher accuracy. Approximately 22% of attempts were identified as fast and inaccurate by this classifier and were classified as non-effective activity use. These attempts were a subset of the attempts identified by the previous classifier, i.e., the low accuracy subset.

3.5 Effective activity use models

We included the counts of "effective activity use" from each classifier in two different models and compared the models with the baseline model. These analyses tell us whether, when considering effective activity use, we are able to capture the learning benefit of engaging with activities.

3.5.1 Active learning use of activities as identified by the duration-based classifier

When using the duration-based classifier, we found that the number of effective activity use was not related to quiz performance, $\beta = -0.33$, $p = .101$ and this model did not improve fit to the data, $\chi^2 = 2.69$, $p = .103$ (see Table 2).

3.5.2 Active learning use of activities as identified by the duration+accuracy classifier

When we considered the counts obtained using the duration+accuracy classifier, we found that greater effective activity use predicted better quiz performance, $\beta = 0.17$, $t\,(3053) = 2.56$, $p = .011$, and this effect was 1.2 times larger than that of accessing more eText pages, $\beta = 0.14$. This model provided a better fit to the data, $\chi^2 = 6.63$, $p = .010$ (see Table 2).

4. DISCUSSION

The two main aims of this study were (1) to investigate the relative benefits of completing activities versus reading in an online course in which completing activities was mandatory, and (2) to explore the key features of effective active engagement with activities and how to detect them in student online behavior.

Previous research suggests that the most beneficial practice activities involve effortful, active engagement and knowledge manipulation by the students [9,18,19]. Indeed, we found evidence that features connected with effort and engagement with the activities were better predictors of learning than completing activities per se (time spent and accuracy). However, overall, we found that, when activities are required and graded, completing more activities is not necessarily a good predictor of improved learning. Instead, spending more time completing the activities and being more accurate across attempts, are better predictors of improved quiz performance. These analyses offer the perfect case-study for the "doer effect" and the characteristics of the learning activities that contribute to improved learning outcomes.

Across all models, more reading (accessing more eText pages) remained the best overall predictor of learning outcomes, even when compared to features indicative of active engagement with the activities. There are multiple reasons for this finding. It is possible that students who accessed the eText were engaging in active learning by autonomously searching answers to activities. Indeed, in a departure from previous studies [13], the activities in this course offered only corrective feedback, implicitly encouraging students to seek more information in the eText, which might have contributed to the results presented here. Another possibility is that better students, who ultimately perform better in the course, access a course material that is not mandatory or rewarded. The reduced correlation between pages read after starting the quiz and quiz performance, suggest that this possibility of a third variable explanation is somewhat less likely than the first possibility that reading behavior in this course is associated with active learning of completing the activities because of the type of feedback used in the activities.

Another main novelty of the present work is the development of analytical processes to identify which activity engagement might be productive and which might not. Under the assumption that the same activity might be completed effortfully and involve knowledge manipulation or only involve "action", we developed two classifiers. The first classifier took into account only the duration of the attempt, whereas the second classifier took into account the duration as well as accuracy of the attempt. The outcome of the first detector did not seem to improve the model fits predicting quiz performance. Conversely when we tested activity use considering only attempts that were classified as effective active learning by the second classifier, we saw that greater effective activity use was a positive predictor of better quiz performance. In fact, greater effective activity use as defined by the second classifier resulted in 1.2 times better quiz performance than accessing more eText pages. Conversely, considering every activity attempt was a negative predictor of quiz performance.

The difference in outcomes between the two classifiers suggests that time to solve a problem by itself might not be sufficient to identify gaming. Fast but accurate attempts might be effective or at least do not negatively impact performance. One possibility is that students learn from fast correct attempts or that fast and correct attempts reflect already learned knowledge. This finding is also congruent with previous findings that some students or some activities might not be harmed by gaming [1,17].

Our approach to defining classifiers differs from previous approaches in educational data mining. We used an explanatory approach; our gaming classifiers were very simple and identified gaming events based on initial data analytics and the literature. This approach might yield less predictive models than previous efforts using more complex (and potentially more predictive) models [1]. However, one benefit of our approach is its explanatory power. The gaming detectors we created can not only identify gaming behavior from the student data but also contribute to a better understanding of what characterizes these types of behaviors (see also [17]).

The findings presented here are a critical first step towards developing effective active learning activities in online courses. Given the greater student autonomy often associated with online courses, it is important to develop methods to identify effective activity use. Critical next steps are to create generalized detectors that can be used online to provide students with feedback not only about the content of the activity, but also their use of the activity as active learning tool. For example, the activity could alert the student to the fast pace and low accuracy and suggest that they try a different approach to the task. Similar classifiers of these "gaming the system" behaviors have been suggested before in the context of intelligent tutoring systems with good success [3].

In sum, the work presented here suggests that not all activity use is active learning and therefore contributes to better learning outcomes. Some activity use might reflect "gaming the system" behaviors that might yield high immediate scores but are not reflective of better learning and later quiz performance (for a discussion see [5]). Similarly, not all reading is passive learning, and intentional use of reading materials might reflect active learning. Accordingly, it is important to be able to detect when students are engaging in active learning and when they are not, regardless of the type of learning activity. The current work establishes an initial step in that direction by identifying which features are associated with active learning engagement when students' complete activities in online courses, and by developing classifiers of this type of behavior that can be adapted and generalized to other courses.

5. ACKNOWLEDGMENTS

This work was supported in part by a National Science Foundation grant (ACI-1443068) toward the creation of LearnSphere.org and by funding from Google to K.R.K.

6. REFERENCES

1. Ryan S. J. d. Baker, Albert T. Corbett, Ido Roll, and Kenneth R. Koedinger. 2008. Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction* 18, 3: 287–314.
2. Ryan SJ d Baker, Albert T. Corbett, Kenneth R. Koedinger, Shelley Evenson, Ido Roll, Angela Z. Wagner, Meghan Naim, Jay Raspat, Daniel J. Baker, and Joseph E. Beck. 2006. Adapting to when students game an intelligent tutoring system. In *International Conference on Intelligent Tutoring Systems*, 392–401.
3. Ryan Baker, Jason Walonoski, Neil Heffernan, Ido Roll, Albert Corbett, and Kenneth Koedinger. 2008. Why Students Engage in "Gaming the System" Behavior in Interactive Learning Environments. *Journal of Interactive Learning Research; Charlottesville* 19, 2: 185–224.
4. Lucy Barnard, William Y. Lan, Yen M. To, Valerie Osland Paton, and Shu-Ling Lai. 2009. Measuring self-regulation in online and blended learning environments. *The Internet and Higher Education* 12, 1: 1–6. https://doi.org/10/b6sf5z
5. Robert A. Bjork, John Dunlosky, and Nate Kornell. 2013. Self-Regulated Learning: Beliefs, Techniques, and Illusions. *Annual Review of Psychology* 64, 1: 417–444.
6. Paulo F. Carvalho, David W. Braithwaite, Joshua R. de Leeuw, Benjamin A. Motz, and Robert L. Goldstone. 2016. An In Vivo Study of Self-Regulated Study Sequencing in Introductory Psychology Courses. *PLOS ONE* 11, 3: e0152115–e0152115.
7. Paulo F. Carvalho, Elizabeth A. McLaughlin, and Kenneth R. Koedinger. 2017. Is there an explicit learning bias? Students beliefs, behaviors and learning outcomes. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, 204–209.
8. Donald S. Ciccone and John W. Brelsford. 1976. Spacing repetitions in paired-associate learning: Experimenter versus subject control. *Journal of Experimental Psychology: Human Learning and Memory* 2, 4: 446–455.
9. L. Deslauriers, E. Schelew, and C. Wieman. 2011. Improved Learning in a Large-Enrollment Physics Class. *Science* 332, 6031: 862–864. https://doi.org/10.1126/science.1201783
10. Jonathan Fernandez and Eric Jamet. 2017. Extending the testing effect to self-regulated learning. *Metacognition and Learning* 12, 2: 131–156. https://doi.org/10/gbms77
11. S. Freeman, S. L. Eddy, M. McDonough, M. K. Smith, N. Okoroafor, H. Jordt, and M. P. Wenderoth. 2014. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences* 111, 23: 8410–8415.
12. René F. Kizilcec and Geoffrey L. Cohen. 2017. Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. *Proceedings of the National Academy of Sciences* 114, 17: 4348–4353.
13. Kenneth R. Koedinger, Jihee Kim, Julianna Zhuxin Jia, Elizabeth A. McLaughlin, and Norman L. Bier. 2015. Learning is not a spectator sport: doing is better than watching for learning from a MOOC. *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15*: 111–120.
14. Kenneth R. Koedinger, Elizabeth A. McLaughlin, Julianna Zhuxin Jia, and Norman L. Bier. 2016. Is the doer effect a causal relationship? *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*: 388–397. https://doi.org/10.1145/2883851.2883957
15. Norbert Michel, J J Cater, and Otmar Varela. 2009. Active versus passive teaching styles: An empirical study of student learning outcomes. *Human Resource Development* ... 20, 4: 397–418. https://doi.org/10.1002/hrdq
16. Kayla Morehead, Matthew G Rhodes, and Sarah DeLozier. 2016. Instructor and student knowledge of study strategies. *Memory* 24, 2: 257–271.
17. Kasia Muldner, Winslow Burleson, Brett Van de Sande, and Kurt VanLehn. 2011. An analysis of students' gaming behaviors in an intelligent tutoring system: predictors and impacts. *User Modeling and User-Adapted Interaction* 21, 1–2: 99–135. https://doi.org/10/bvkpzd
18. Henry L. Roediger and Jeffrey D. Karpicke. 2006. The Power of Testing Memory: Basic Research and Implications for Educational Practice. *Perspectives on Psychological Science* 1, 3: 181–210.
19. Carl E Wieman. 2014. Large-scale comparison of science teaching methods sends clear message. *Proceedings of the National Academy of Sciences of the United States of America* 111, 23: 8319–8320.