A tale of two guessing strategies: Interpreting the time students spend solving problems through online log data

Matthew W. Guthrie, Tom Zhang, and Zhongzhou Chen
Department of Physics, University of Central Florida, Orlando, FL, 32816

Interpretation of student behavior in online learning platforms based on log data is complicated by not being able to directly observe the learner. In this paper, we attempt to identify data patterns that signal either guessing on assessment problems or disengaging from the task for students while working through homework modules in an introductory physics class by contrasting data from the general student population with those who completed homework modules in controlled, observed environments. We found that abnormally short problem solving attempts that were previously modeled as a single guessing or answer copying behavior actually consisted of two different types of "guessing" behavior: rapid and strategic guessing. Both types were associated with lower levels of self-confidence, but had different distribution among proctored and unproctored student populations. More importantly, the fraction of rapid guessing increased significantly after campus closure due to COVID-19, but the fraction of strategic guessing remained constant.

2020 PERC Proceedings edited by Wolf, Bennett, and Frank; Peer-reviewed, doi.org/10.1119/perc.2020.pr.Guthrie

I. INTRODUCTION

Time-stamped log data from online learning platforms provide rich information about students' learning behavior [1–3]. However, one significant challenge in the mining of log data is to reliably relate the patterns detected to actual student behavior, especially when the data are noisy. For example, if a student submitted an answer to a problem shortly after opening it, how short does the duration need to be for researchers to conclude that the student generating those events was guessing [4, 5]? Or, if no new log event is recorded over for a period time, how long does that period need to be before researchers can determine that the student has disengaged with the learning process [6]? While the answers to these questions can be estimated through data analysis techniques alone [7, 8], a more reliable method is to recruit and observe students in a lab environment, and correlate the log data from those students with their observed learning behavior, as exemplified in Baker's studies of student disengagement [9].

In the current study, we adopt a simpler protocol similar in principle to Baker's approach to understand students' log data during online problem solving. Our goal is to identify patterns in the data that are indicative of students being completely disengaged from the learning materials, such as walking away or quickly clicking through the material. Therefore, we recruited students to complete parts of their online homework in a in a regular classroom, with one proctor who took attendance but did not observe students' learning behavior. No additional requirements were imposed on the students' behavior to allow for the collection of log data that more closely corresponds to students' "natural" problem solving processes, which could involve temporary disengagement from tasks or guessing on assessment problems. We hypothesized that by contrasting proctored students' data with the log data collected from the rest of the class, we would be able to identify patterns in the data that correspond to complete disengagement from tasks.

Additionally, on some assessment problems, we asked students to rate their confidence level in their responses to the problems. These data can assist in distinguishing between a guessing attempt from a quick answer to a simple question.

A practical application for identifying the abnormal signals associated with disengagement in student data is to monitor the change in their level of engagement throughout the semester. This application is especially valuable and timely amidst the COVID-19 outbreak, as instructors and researchers try to quantify the impact of the abrupt shift to distant learning on students' learning behavior.

In this paper, we first present our analysis and comparison of both the time-stamped log data and the survey data between the proctored and unproctored student samples. The comparison suggests that what had previously been treated as a single type of "guessing" or "copying" behavior [10, 11] may actually consist of two different types of behavior: "rapid" guessing during which students barely read the problem text, and "strategic" guessing, in which students likely read the problem but did not fully solve the problem. More specifically, we will answer the following research questions:

RQ 1 What are reasonable cutoffs for exceptionally brief and exceptionally long assessment attempts?

RQ 2 How are students' estimates of their confidence on assessments related to their attempt durations?

RQ 3 How does the fraction of brief assessment attempts change over the semester, in particular after the COVID-19 outbreak?

II. STUDY DESIGN AND METHODS

The log data analyzed in this study were generated from students' interactions with online learning modules (OLMs), implemented in the Obojobo Next online platform [12] developed at the University of Central Florida Center for Distributed Learning. The OLMs were assigned as homework in a Spring 2020 introductory calculus-based mechanics course at the University of Central Florida. Each module contains an instructional component with learning materials and an assessment component which consists of one or two multiple choice problems assessing students' understanding of the content presented in the learning materials. The assessments contain mainly problem-solving and calculation questions. The individual modules were designed for students to complete in 20-30 minutes, and are organized in a mastery learning format as detailed in Ref. [13]. Each student was allowed 5 attempts on each assessment and required at least one attempt on the assessment before being given access to the instructional material. For each of the first three assessment attempts in every module, students received a new isomorphic problem set, whereas, on the 4th and 5th attempt, the problem set on the 1st and 2nd attempts were repeated. Students received a 10% penalty if they passed the module on their 4th or 5th attempt.

A. Implementation of proctored sessions

Students were invited to participate in the study by completing parts of their homework assignments in a proctored environment. Students received 1% of extra course credit for participating in a homework session. Sessions were announced through two short statements before regular lecture and two emails to the class. Students were permitted to attend a maximum of 2 out of 9 available sessions. 46 out of the 273 students enrolled in the course participated in the observations, and we recorded 61 individual sessions. Each session lasted between 1 and 2 hours [14], and students were required to stay for at least 1 hour to receive the extra credit incentive.

The proctors of the homework sessions refrained from interacting with participants as much as possible to minimize the impact of their presence on student's behavior. Proctors

only spoke with participants during a short appeal at the beginning of the session asking students to work on the homework modules during the session, as well as recording their name, student ID number, and sign-in/sign-out times.

B. Survey questions

For 43 out of the 65 OLMs, the assessment component also contained a survey question asking students to rate their confidence of their answer to the assessment problems on a 5-point Likert scale: "Strongly/Very Unconfident," "Unconfident," "Neutral," "Confident," "Strongly/Very Confident." Following common practice [15, 16], we collapsed students' responses into three categories: "Confident," "Neutral," and "Unconfident" in data analysis.

For each of the first three assessment attempts in every module, students were presented with isomorphic problem sets related to the topic of the module. After the 3rd attempt, students received the 1st and 2nd sets of problems again and were assigned a lower score if they subsequently passed the module. In the current paper, we only analyze students' 1st, 2nd, and 3rd attempts for each module, as guessing behavior significantly changes when students are presented with a problem they have already seen. We define the duration of an attempt to be the amount of time from opening an assessment to submitting a response.

III. RESULTS

A. Assessment attempt duration analysis

To answer **RQ 1**, we plot the density distribution of attempt durations on a log scale, for attempts conducted by students in the proctored session and attempts conducted by all other students before and after the campus closure due to the COVID-19 outbreak in Fig. 1A. Attempt durations are cut off at 10^4 seconds since only 2.1% of all attempts were longer. The longest attempt recorded in the proctored session was 3,707 seconds. The fraction of attempts between 10^2 and 10^3 seconds is much higher for students in the proctored sessions than in the two other distributions, whereas the trend is reversed for attempts below 35 seconds. Additionally, Kruskal-Wallis H and Mann-Whitney U tests show the before closure, after closure, and proctored distributions are all significantly different from one another (see Table I) [17].

For all three populations, attempts shorter than 35 seconds formed its own cluster that is separated from the main distribution when plotted in log scale. For the students in the proctored session, the peak of the distribution lies between 15 and 35 seconds, while for attempts after the COVID-19 outbreak a prominent peak lies under 15 seconds. For attempts made before COVID-19, there is no obvious peak.

We plot in Fig. 1B the duration of attempts under 100 seconds on a linear scale, to examine the distribution of brief attempt durations in detail. A very small fraction of attempts in Fig. 1B from the proctored students were less than

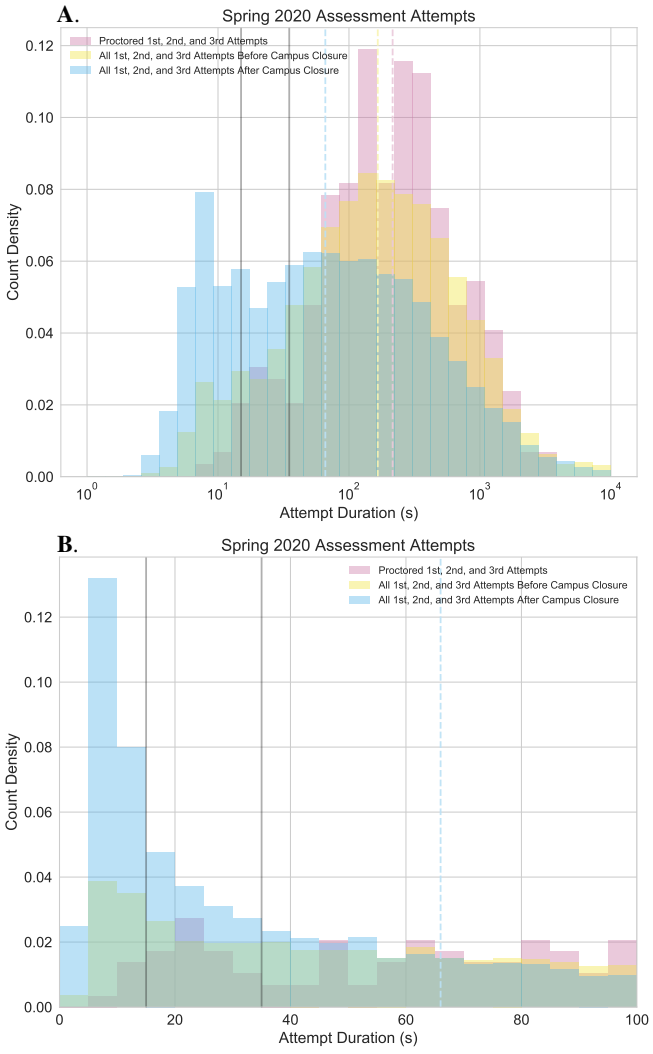

FIG. 1. Normalized histograms for three groups of students: those who participated in proctored homework sessions, for all attempts before campus closure, and attempts after campus closure. Dashed vertical lines indicate the median time for each group and solid vertical lines indicate 15 and 35 seconds. Panels **A** and **B** show the same data on logarithmic and linear scales.

15 seconds, while the other two distributions have significantly higher fractions, especially for attempts made after the COVID-19 outbreak. The brief attempts for proctored students peaked between the range of 15 to 35 seconds, within which the difference between the three distributions are relatively smaller.

A number of previous studies attributed very short attempts to guessing or answer copying behavior since the duration is barely long enough for the student to finish reading the problem text [4, 5, 10]. Based on the current results, we hypothesized that those brief attempts may consist of two different types of guessing behavior: "rapid" guessing occurs mostly under 15 seconds, is rarely observed among students in the proctored sessions but is more frequent among other students, especially after the COVID-19 outbreak. Analyzing

TABLE I. Kruskal-Wallis H and Mann-Whitney U statistical tests comparing each group of attempt durations shown in Fig. 1.

Group (median)	Comparison	Statistic	p-value
Proctored (216 s)	Before Closure	$H = 5.38$	$p = 0.02$
Before Closure (166 s)	After Closure	$U = 1.24\times10^{6}$	$p < 0.01$
After Closure (66 s)	Proctored	$U = 1.13\times10^{8}$	$p < 0.01$

the survey data in Sec. III B allows us to further investigate this hypothesis. "Strategic" guessing lasts between 15 and 35 seconds and is observed at a similar frequency among proctored and unproctored students. While the duration is still much shorter than necessary for answering most of the assessment problems, it is likely long enough for students to at least quickly read the text of the problem. The 26 rapid and strategic guesses submitted by proctored students were distributed across 12 different modules.

B. Survey data

To answer **RQ 2**, we plot the distributions of each category of confidence: confident, neutral, unconfident, and no response in Fig. 2, grouped by the attempt duration with a maximum duration of 150 seconds. As shown in Fig. 2, for the unproctored population, the fraction of confident responses is much lower than neutral and unconfident responses in attempts less than 15 seconds. This difference is still evident for responses between 15 and 30 seconds, although it is much smaller. For attempts greater than 30 seconds, the difference either disappeared or is reversed. For the proctored session, the highest count of neutral response occurred for attempts between 15 and 30 seconds, whereas the few unconfident responses are distributed evenly across each bin.

C. Variation in student behavior after campus closure due to COVID-19

To explore the impact of the COVID-19 outbreak on students' online learning behavior (**RQ 3**), we plot in Fig. 3 the fraction of either rapid or calculated guessing in all attempts on every module assigned in the Spring 2020 semester.

The two vertical lines in Fig. 3 separate modules before and after campus closure according to either their release or due dates. Modules to the left of the dashed line were due before closure, and modules to the right were due after the closure. The solid line separates modules that were released before/after the campus closure. Linear regressions for both guessing strategies over either the entire semester or periods separated by the campus closure are listed in Table II.

The fraction of both types of guessing increased slightly at nearly identical rates over time prior to campus closure, with occasional spikes on certain modules. In contrast, the fraction of rapid guessing jumped to over 40% for most of the modules released after campus closure, while the fraction of strategic guessing remained largely unchanged.

TABLE II. Linear fit parameters for the distributions in Fig. 3. Rapid guessing durations are shorter than 15 seconds; strategic guessing attempts are between 15 and 35 seconds in duration.

Guessing group	Slope	Intercept	R	p-value
Rapid (full sequence)	0.0053	−0.012	0.80	$p < 0.001$
Strategic (full sequence)	0.0012	0.080	0.49	$p < 0.001$
Rapid (before closure)	0.0022	0.004	0.52	$p < 0.001$
Strategic (before closure)	0.0021	0.065	0.45	$p = 0.003$
Rapid (after closure)	0.0110	−0.330	0.77	$p < 0.001$
Strategic (after closure)	0.0014	0.068	0.35	$p = 0.067$

IV. DISCUSSION AND FUTURE WORK

In this paper, we present multiple pieces of evidence suggesting that abnormally short problem solving attempts in an online environment could stem from two distinct types of guessing behavior: "rapid" guessing, which is less than 15 seconds in duration, and "strategic" guessing, which takes place in roughly 15 to 35 seconds. To the best of our knowledge, this distinction has never been made in previous papers studying students' brief problem solving behavior in online environments [4, 5].

The distinction between the two types is supported by three findings:

1. Rapid guessing is rarely observed among students in the proctored homework session but observed much more frequently among students completing assignments on their own. On the other hand, strategic guessing is observed with similar frequency in both populations.
2. While both types of guessing are associated with a reduced level of self-confidence, students are less confident about their answers when conducting rapid guessing compared to those who are conducting strategic guessing.
3. The fraction of rapid guessing dramatically increased shortly after the COVID-19 related campus closure, while the fraction of strategic guessing remained the same.

These observations also shed light on the different behavioral nature of the two types of guessing. Rapid guessing likely occurs when the student is disengaged from learning resources and submits an answer without reading or having only briefly read the problem text. It is also possible that a fraction of those submissions stem from students copying their answers from a peer. The fraction of rapid guessing can thus serve as a detector of disengagement among students. In our current analysis, the significant increase in rapid guessing shows the profound impact of the abrupt shift to distance learning on students' ability to continue with their course work.

On the other hand, strategic guessing is likely being performed by students who are engaged in the learning process but do not initially know how to solve the problem. Although

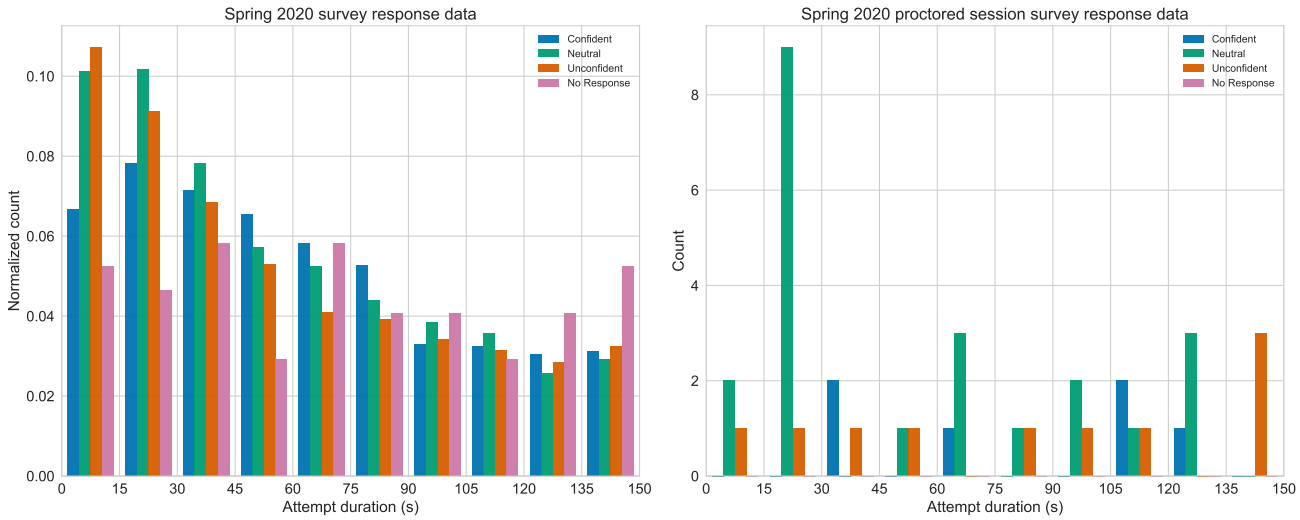

FIG. 2. Survey responses for attempts with duration less than 150 seconds. Each category was normalized independently to highlight the overall distribution of each category of survey responses.

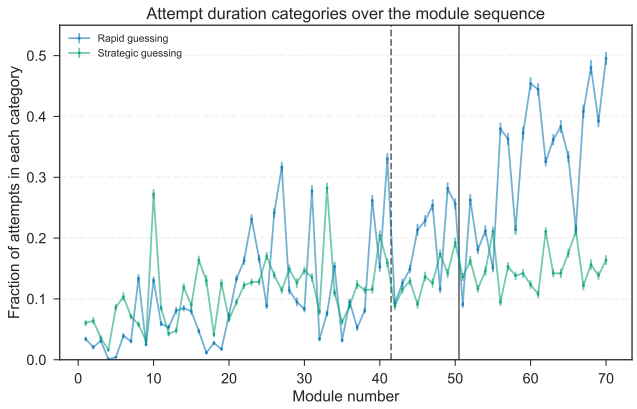

FIG. 3. As the semester progressed, students changed their assessment strategies. The fraction of students who followed "strategic" guessing behavior was approximately constant, while the fraction of students who engaged in a "rapid" guessing strategy sharply increased after the campus-wide transition to online-only instruction. The dashed vertical line marks the time at which modules were released before spring break but due after spring break and the solid vertical line marks the modules released and due after spring break.

we used "guessing" in the name, it could also originate from students who either made an educated guess or misinterpreted the problem and thought that it could be solved very quickly. This could explain why more strategic guessing students rated "confident" on the survey question. A more appropriate name may be given to this type of behavior in follow-up studies. Remarkably, the fraction of strategic guessing remains roughly the same after the campus closure. A possible explanation is that this type of behavior may be a common study strategy adopted by students in a mastery learning setting.

While all three pieces of evidence point to two different types of guessing, an extensive amount of follow-up data analysis will be needed to confirm the existence of the two types, determine more accurate duration cutoffs for each type, and further reveal the cognitive and metacognitive process behind each type. Of particular importance is to examine how problem solving behavior depends on the context, type, and difficulty of the problem, and on what fraction of the problems we can detect the two types of guessing. While rapid guessing is likely context-independent as students have not had time to interact with the material, strategic guessing may only occur on certain types of problems. Furthermore, to what extent does the design of the learning experience and online platform impact students' problem solving behavior and strategy? Finally, the percentage of correct answers for each type of guessing needs to be carefully examined, which will provide insight into the behavioral nature of each type of guessing behavior.

In this paper, we have demonstrated that new insights into students' online learning behavior can be obtained by comparing the log files of all students to those who interacted with the resources in a proctored environment, even with minimum or no interaction with the proctor. A potentially fruitful future direction is to see if the same technique can be applied to the analysis of other types of data, such as problem solving attempts with normal or abnormally long duration, the duration of learning from instructional resources, or interacting with other online learning systems.

ACKNOWLEDGMENTS

We would like to thank the Learning Systems and Technology team at UCF for developing the Obojobo platform. This research is partly supported by NSF Award No. DUE-1845436.

[1] J. Park, K. Denaro, F. Rodriguez, P. Smyth, and M. Warschauer, Detecting changes in student behavior from clickstream data, in *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, LAK '17 (Association for Computing Machinery, New York, NY, USA, 2017) pp. 21–30.

[2] R. S. Baker and P. S. Inventado, Educational data mining and learning analytics, in *Learning Analytics* (Springer New York, 2014) pp. 61–75.

[3] C. Dede, ed., *Data-Intensive Research in Education: Current Work and Next Steps* (Computing Research Association, Arlington, VA, 2015).

[4] D. J. Palazzo, Y.-J. Lee, R. Warnakulasooriya, and D. E. Pritchard, Patterns, correlates, and reduction of homework copying, Physical Review Special Topics - Physics Education Research **6**, 10.1103/PhysRevSTPER.6.010104 (2010).

[5] R. Warnakulasooriya, D. J. Palazzo, and D. E. Pritchard, Time to completion of web-based physics problems with tutoring, Journal of the Experimental Analysis of Behavior **88**, 103 (2007).

[6] V. Kovanović, D. Gašević, S. Dawson, S. Joksimović, R. S. Baker, and M. Hatala, Penetrating the black box of time-on-task estimation, in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge - LAK'15* (ACM Press, 2015).

[7] R. S. Baker, A. T. Corbett, and K. R. Koedinger, Detecting student misuse of intelligent tutoring systems, in *Intelligent Tutoring Systems* (Springer Berlin Heidelberg, 2004) pp. 531–540.

[8] R. S. Baker, A. T. Corbett, K. R. Koedinger, and A. Z. Wagner, Off-task behavior in the cognitive tutor classroom, in *Proceedings of the 2004 conference on Human factors in computing systems - CHI'04* (ACM Press, 2004).

[9] R. S. Baker and L. M. Rossi, Assessing the disengaged behaviors of learners, Design recommendations for intelligent tutoring systems **1**, 153 (2013).

[10] Z. Chen, M. Xu, G. Garrido, and M. W. Guthrie, Relationship between students' online learning behavior and course performance: What contextual information matters?, Phys. Rev. Phys. Educ. Res. **16**, 010138 (2020).

[11] Z. Chen, G. Garrido, Z. Berry, I. Turgeon, and F. Yonekura, Designing online learning modules to conduct pre- and post-testing at high frequency, in *Physics Education Research Conference 2017*, PER Conference (Physics Education Research Topical Group and the American Association of Physics Teachers, Cincinnati, OH, 2017) pp. 84–87.

[12] Z. Berry, I. Turgeon, and F. Yonekura, *Obojobo Next* (2020).

[13] M. W. Guthrie and Z. Chen, Comparing student behavior in mastery and conventional style online physics homework, in *Physics Education Research Conference 2019*, PER Conference (Provo, UT, 2019).

[14] The variation in session length was due to constraints in room availability.

[15] M. S. Matell and J. Jacoby, Is there an optimal number of alternatives for likert scale items? Study I: Reliability and validity, Educational and Psychological Measurement **31**, 657 (1971), https://doi.org/10.1177/001316447103100307.

[16] N. Silver, How We're Tracking Donald Trump's Approval Ratings (2017).

[17] Note: we use the Kruskal-Wallis test to compare the proctored and before closure distributions because all of the proctored sessions occurred before campus closure and are therefore included in the before closure subset of attempt data.