Comparing student behavior in mastery and conventional style online physics homework

Matthew W. Guthrie[1] and Zhongzhou Chen[1]

[1]*Department of Physics, University of Central Florida, Orlando, FL, 32816*

We compared students' learning behavior when completing identical online calculus-based physics homework assignments organized in two ways. One was designed for mastery learning where content is divided into smaller units, and students are required to attempt the assessment once before accessing the content. Students can proceed to the next unit after passing the assessment either before or after studying the content. The second is a conventional design in which students first study a set of instructional materials equivalent to several mastery units then complete multiple assessment problems at once. Our major findings are: 1. in the mastery condition, students solved more problems correctly either immediately after studying the instructional content, or on attempts before accessing the instructional content; 2. for students who solved similar numbers of problems correctly, the mastery condition students spent significantly less time studying compared to the traditional condition students; and 3. students who did not pass mastery units on their initial assessment attempts spent similar amounts of time studying as traditional condition students.

I. INTRODUCTION

Almost every conventional physics textbook is organized by chapters, with each chapter covering all of the concepts related to a general topic such as mechanical energy or linear momentum, and a considerable number of homework problems appended to the end of each chapter. In a traditional class, students' mastery of content knowledge is assessed by either a quiz or an exam after finishing one or several chapters. While recent development of online learning platforms enables instructional content to be easily disseminated through the Internet, most online courses are still organized in the traditional "instruction - homework - quiz" style [1, 2].

An alternative to this conventional style of instruction, mastery learning, as proposed by Bloom [3] and Keller [4], has been adopted by instructors across multiple disciplines [5]. In mastery learning, the conventional chapters are broken down into a sequence of smaller units, where each unit focuses on developing a single skill or concept. After initial instruction, students who could demonstrate mastery of the concept or skill on the assessment can advance to the next unit, while those who could not are allowed more time to study and are provided with additional scaffolding until mastery is achieved. By enabling students to spend different amounts of time and effort on learning, it is proposed that significantly more students in a mastery learning condition will achieve the desired level of mastery. There have been several recent attempts at implementing mastery style online homework in introductory physics courses with encouraging outcomes [4, 6, 7].

To what extent does student learning behavior differ when presented with the same instructional content organized either in a traditional textbook style or a mastery learning format? In particular, do introductory physics students allocate different amounts of time to achieve mastery in a mastery learning condition, as intended by Bloom and Keller? Advancements in online learning technology enable researchers to answer these questions by providing detailed information on student learning behavior via clickstream data. The current study explores the answer to these questions by organizing the exact same set of online instructional materials into two different versions, one according to the principles of mastery learning, and the other resembling a conventional textbook experience as much as possible. These instructional materials (text, practice problems, and assessment problems) were assigned as homework to two sections of the same introductory physics course taught in a traditional lecture format. By comparing the students' learning behavior under the mastery and traditional conditions, we test the following three hypotheses developed directly from the principles and hypotheses of mastery learning that more students will achieve mastery by spending different amounts of time on learning:

H1 More students will answer more assessment problems correctly in the mastery condition than in the traditional condition.

H2 For students who answered similar numbers of assessment problems correctly, the time spent on learning will be more uniform in the traditional condition, and more diverse in the mastery condition. In other words, some students in the mastery condition will spend significantly less time while others will spend significantly more time.

H3 In the mastery condition, students who pass more units on their initial assessment attempts spend less time learning than those in the traditional condition, while students who fail more units on initial attempts spend more time.

II. STUDY DESIGN AND METHODS

We employed a flipped online AB experiment design similar to two earlier studies [8, 9]. Two versions of online homework were created for two chapters of a calculus-based introductory physics course, covering the topics of mechanical energy and linear momentum. The two versions of the mechanical energy homework were each assigned to one of two large sections of the course, both having about 220 students and being taught in a traditional lecture style. Each section was then assigned the other version of the homework for the linear momentum chapter. This design ensures that if similar differences between the two versions were observed for both chapters, they are not due to extraneous differences such as instructor and student population between the two sections. Students were instructed to complete each set of homework over a period of two weeks for course credit, during which the content was covered in lecture. No additional homework was assigned to the students during the four-week portion of the class involved in this study.

A. Creation of Mastery and Traditional Learning Conditions

The mastery style (labeled Mastery throughout the paper) online homework consists of sequences of Online Learning for Mastery (OLM) modules [10–13], implemented on the Obojobo platform developed by the Center for Distributed Learning at the University of Central Florida (UCF) [14]. Each OLM module contains an assessment component (AC) and an instructional component (IC). In the Mastery condition, each module focused on one concept or skill, with the AC of the module containing 1 to 2 assessment problems focused entirely on the skills being assessed in the AC. The IC of each module contains 1 to 3 instructional pages with text, figures, and practice problems that provide immediate feedback and do not count toward the final grade. Each student was allowed 5 attempts on the AC but required 1 attempt before being allowed to access the IC. The justification for this requirement is that most students were introduced to the contents either during lecture or in high school courses, which serves as initial instruction in the original mastery learning design. To proceed to the next module, a student must either correctly answer all questions on the AC during any of the attempts or use up all 5 attempts. If a student passes the AC on the initial attempt before accessing the IC, they are encouraged to skip the IC and directly proceed to the next module.

The mechanical energy chapter contains 10 OLM modules, while the linear momentum chapter contains 9 modules [15].

The traditional style (labeled Traditional) homework was created by re-organizing the exact same contents of the OLM modules. Each sequence of Mastery modules was combined into three Traditional modules, with each Traditional module containing the same content of 3 - 4 Mastery modules, including 3 - 6 pages and 3 - 7 AC problems. To minimize extraneous differences, the Traditional modules were also delivered via the Obojobo platform. The most significant difference was that students were first presented with the IC and not required to make an initial attempt on the AC. They were also given 5 attempts on the AC and must either answer all problems correctly or use up all the attempts, in which case the score of their highest attempt was recorded to the gradebook. In both conditions, a student received an isomorphic set of AC problems on their first 3 attempts, after which the 1st and 2nd set was presented for the 4th and 5th attempts. If a student fails the AC on the initial attempt and passes after studying the IC, we can attribute the knowledge gain solely to the IC. The symbolic solutions to the first set of problems were given as practice problems in the instruction but the final number for the answer was not given.

B. Measuring Learning Duration and Correct AC Responses

Under both conditions, most students visited the IC in a given module a single time. In 4% of the cases in the Mastery condition and 8% of cases in the Traditional condition, a student made a second visit to the IC of a module that was at least 50% as long as the longest visit and longer than 60 seconds. In those cases, the two visits were considered as a single visit and the attempt in between was neglected from the analysis. For the even rarer case of a student making 3 or more visits to the IC that are similar in duration, only the two longest visits were combined. We refer to either the single visit or the combined visit as a student's Major Study Session (MSS) for that module, and the duration of a MSS is used as the measure of students' learning time on that module.

In this study, we use the number of AC problems answered correctly either before or immediately after studying the IC as a proxy for a student's level of knowledge mastery. Specifically, we count the number of correctly answered problems on both the first attempt after the MSS and the first or second attempt before the MSS, if they exist. Students in the Traditional condition were grouped based on the number of correct AC responses on each module and were compared to students in the Mastery condition who correctly answered similar numbers of AC problems on the corresponding Mastery modules.

III. RESULTS

To examine **H1** - **H3**, we compare the number of correctly answered assessment problems (N_p) before or immediately after a student studies the IC, and the duration of students' MSS (T_s), for each Traditional module. If a student was in the Mastery condition, their N_p and T_s is the sum of N_p and

TABLE I. Mann-Whitney U comparisons of N_p distributions for Mastery and Traditional groups, including the mean N_p for students in each condition. Asterisks indicate significance at $\alpha = 0.05$ levels.

Module	Mastery $\overline{N}_p$	Traditional $\overline{N}_p$	U statistic	p-value
1	5.36	5.19	19340.5	0.03^*
2	3.28	3.04	17835.5	0.03^*
3	1.80	1.28	13051.5	$< 0.01^*$
4	4.59	4.11	14512.0	$< 0.01^*$
5	2.42	2.12	10997.0	0.01^*
6	3.25	3.05	7377.0	0.11

T_s from the modules that were combined into a corresponding Traditional module.

For **H1**, we compared the number of students with different amounts of N_p on each Traditional module in both conditions. In Figure 1, we grouped students from low to high into three tiers of N_p for each module, representing the size of each group using the area of a dot. The N_p groupings were chosen such that each group on each module contains roughly 1/3 of the students. For all six modules, we consistently observe that more students in the Mastery condition have higher N_p than their Traditional peers. Mann-Whitney U (MWU) tests on the population distributions of N_p revealed significant differences between the two conditions in all but the final module, as shown in Table I.

For **H2**, we first plotted the distribution of T_s on a log scale for each N_p group on each module, an example of which is shown in Figure 2. In all 18 cases, we found that the T_s distribution for students in the Mastery condition lies to the left (lower) of the distribution of students in the Traditional condition (Figure 2). In contrast to the stated **H2**, we did not find a separate group of students in the Mastery condition having a higher T_s than the majority of students in the Traditional condition. We show the mean T_s duration for each group on each module in Figure 1, represented by the y-coordinate of each dot. In most of the cases, the mean T_s of the Traditional condition is significantly higher than that of the Mastery condition. The p-values from the MWU tests between the two conditions for each N_p group in each module (adjusted using Benjamini and Hochberg's false discovery reduction for multiple comparisons [16]) shown in Table II confirm that most of the differences were significant, except for all N_p groups on Module 6 and a few other cases. On Modules 2 and 6, the differences are greater for the low N_p groups, whereas for Modules 3, 4, and 5, the difference is greater for high N_p groups. In addition, we tested if students' T_s is a linear function of the N_p grouping on any module. ANOVA results [17] showed that there is a significant linear relation for the Mastery condition on Modules 1 ($F = 9.8$, $p < 0.01$) and 2 ($F = 6.3$, $p < 0.01$).

For **H3**, we identified a subgroup of students under the Mastery condition who did not pass any of the OLM modules corresponding to a given Traditional module on their first or second attempts before accessing the IC (labeled NoPassin2). If **H3** were true (i.e. being able to pass the AC on attempts
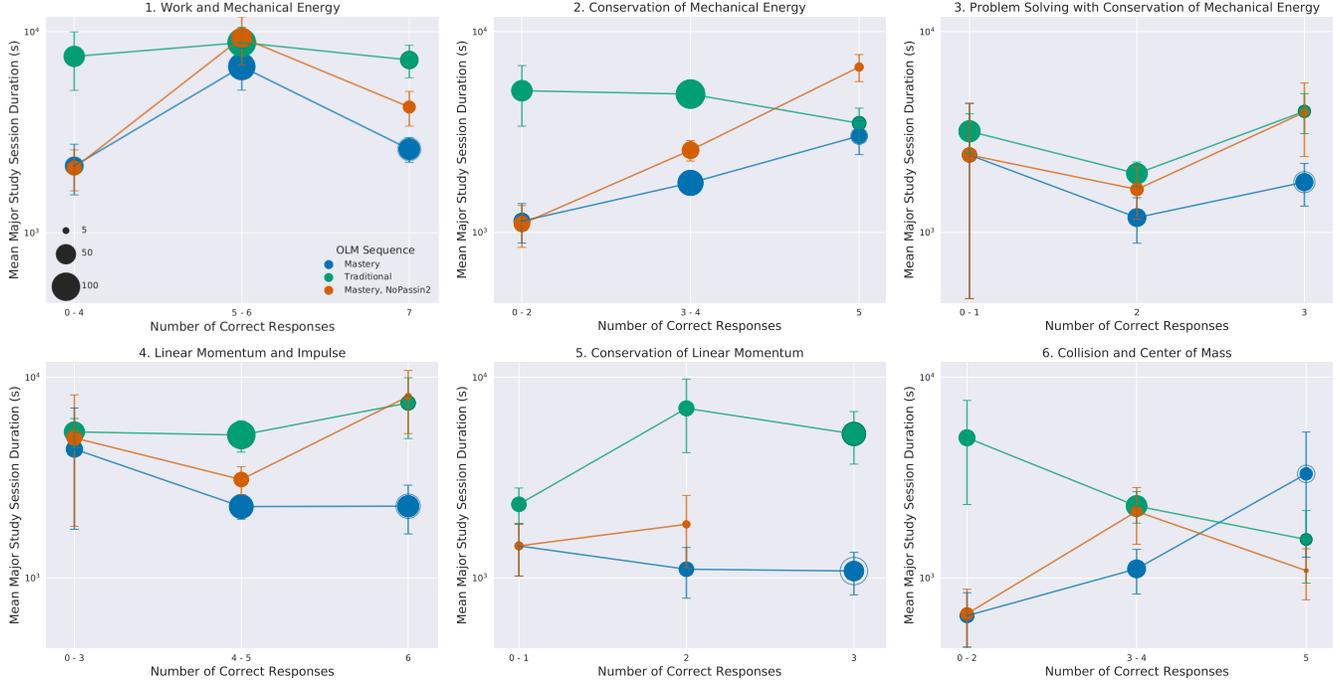
FIG. 1. Students' major study session duration grouped by the number of correct responses in students' attempts after the major study session. Error bars show standard deviations in each distribution of study session durations. Marker size shows the number of students in each group; empty circles show the size of the group if we include students who passed the assessment before studying the IC (and therefore have no MSS duration).
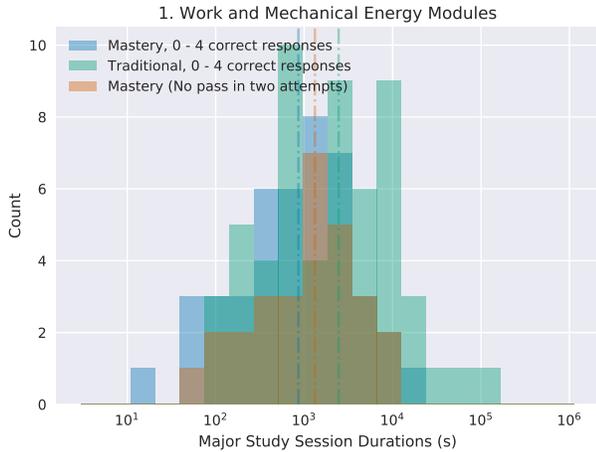

FIG. 2. Major study session durations for students in the mastery and traditional conditions, with a subset of mastery students who did not pass before two attempts.

before studying the IC is the major cause for some students in the Mastery condition to spend less time studying) then the NoPassin2 group would have similar or higher T_s than the Traditional group. As shown in Figure 1 and confirmed by MWU tests in Table II, the NoPassin2 group has similar T_s as the Traditional group (matching N_p) with few excep-

tions. The most significant exception is on Module 2, where both low N_p NoPassin2 groups have significantly lower T_s, and the highest N_p group has higher T_s than the Traditional group. In some cases, such as on Module 5, too few students were in the NoPassin2 group to provide reliable statistics.

IV. DISCUSSION

In this study, we examined **H1**-**H3** by analyzing students' interaction data and comparing N_p and T_s between the two conditions. We found strong evidence to support **H1** by showing that more students in the Mastery condition answered more AC problems correctly on every module. We suggest several possible explanations for this observation. First, most students in the Traditional condition chose to make their first AC attempt after studying the IC, while students in the Mastery condition were forced to make an attempt before accessing the IC. This resulted in students in the Mastery condition taking one more attempt at each AC problem than the majority of students in the Traditional condition. Future analysis will explore whether **H1** is still true when we allow problems answered on the 2nd attempt after study to be counted as correct on the Traditional modules. However, this advantage may be less significant than expected; a separate analysis involving similar OLM modules found that many students consistently submit an answer in a very short amount of time on their initial attempt [11]. Second, the required initial attempt in the Mastery condition may have a Preparation for

TABLE II. Adjusted p-values from Mann-Whitney U comparisons of T_s for students in Mastery and Traditional conditions.

Module	N_p Group	Mastery	NoPassin2
1	0-4	0.01*	0.08
	5-6	0.01*	0.43
	7	< 0.01*	0.17
2	0-2	0.02*	0.02*
	3-4	0.01*	0.03*
	5	0.14	< 0.01*
3	0-1	< 0.01*	< 0.01*
	2	0.04*	0.37
	3	0.02*	0.50
4	0-3	< 0.01*	< 0.01*
	4-5	< 0.01*	0.30
	6	< 0.01*	0.10
5	0-1	0.37	0.37
	2	0.07	0.37
	3	0.05*	—
6	0-2	0.07	0.07
	3-4	0.07	0.39
	5	0.07	0.29

Future Learning (PFL) effect [18], allowing students to focus on information in the IC that is directly related to passing the AC, making learning from the IC an experience similar to Just-in-Time Teaching [19]. Also, it may be more cognitively demanding and emotionally stressful for students in the Traditional condition to correctly solve multiple problems at once, compared to focusing on 1 - 2 problems at a time on each Mastery module.

It must be emphasized that we cannot conclude that students achieved greater mastery of knowledge in the Mastery condition because the assessment problems were not administered under the same setting. A separate common post-test administered in a proctored setting is required to assess the learning outcomes from both conditions.

Our results did not support **H2**, as we found that students in the Traditional condition consistently spent more time on the IC. Remarkably, T_s was not correlated with N_p on any of the modules in the Traditional condition, whereas there was a significant positive correlation for Module 1 and Module 2 in the Mastery condition [20].

We believe that the main reason for the difference in study time, as discussed below in our analysis of **H3**, is that students in the Mastery condition can skip a module if they pass the AC on their initial attempt, which saves a significant amount of time. It may also be possible that the PFL effect mentioned above helped some students to be more focused and efficient in studying the IC. However, there may also be a technical reason for this difference. The IC of a Traditional module is several times longer than an OLM module, making it more likely for students to get distracted or take short breaks when studying the IC. Some of those breaks can be identified by our data processing scripts, but others remain indistinguishable from regular study behavior, contributing to longer T_s for the Traditional condition. While it may seem that students in the Mastery group solved more problems with less study time, it is worth pointing out that, by studying the IC more thoroughly, students in the Traditional condition could have gained knowledge or ability that was not assessed in the AC of the module. Better designed assessments are needed in future studies to assess more implicit gains, such as the ability to transfer [12].

Finally, **H3** is partially supported by the fact that the T_s for the NoPassin2 group is similar or in some cases even higher than that of the Traditional group. This suggests that when students in the Mastery condition did not pass the OLM module on their initial attempt, they tended to spend similar amounts of study time as those in the Traditional condition. One noteworthy exception is on Module 2, where the low N_p group still showed a significant difference between NoPassin2 and Traditional students. In addition, on Module 2, T_s is highly correlated with N_p for students in the Mastery condition, especially for the NoPassin2 group, but not for those in the Traditional condition. Understanding the implication of this exceptional observation requires a substantial amount of future analysis and even follow up study design.

As an exploratory effort in comparing student behavior in an online mastery learning design with a more traditional design, this study provided valuable insight, while leaving several caveats that must be addressed in future studies. First, we focused on a single aspect of student behavior: study time. While study time is at the core of mastery learning, there can be important differences in other behavioral aspects, including how students distribute work over time and how many practice problems they attempt (i.e. the "doer" effect [21]). Also, while most students in the Mastery condition directly skipped the IC of a module after passing on their initial attempt, some did access the IC afterwards. More careful data analysis is needed to account for those after-pass study times, and to examine their impact on the current results.

Alongside the common patterns observed across all six modules, for which we provided several possible explanations, we also reported several unique patterns that are observed only on one or two of the modules. For example, T_s is highly correlated with N_p for the Mastery and NoPassin2 groups only on Module 2. What unique features of Module 2 caused such a strong correlation? Answering this type of question will require conducting similar future studies on significantly more modules covering different topics in physics and other STEM disciplines. Those studies will likely provide valuable new insight into how different design choices for online instructional materials affect students' learning behavior and learning outcomes.

ACKNOWLEDGMENTS

[1] S. Rayyan, C. Fredericks, K. Colvin, A. Liu, R. Teodorescu, A. Barrantes, A. Pawl, D. Seaton, and D. Pritchard, A mooc based on blended pedagogy, J. Comp. Assist. Learn. **32**, 190 (2016).

[2] L. Breslow, D. E. Pritchard, J. DeBoer, G. S. Stump, A. D. Ho, and D. T. Seaton, Studying learning in the worldwide classroom research into edx's first mooc., Research & Practice in Assessment **8**, 13 (2013).

[3] G. Gladding, B. Gutmann, N. Schroeder, and T. Stelzer, Clinical study of student learning using mastery style versus immediate feedback online activities, Phys. Rev. ST Phys. Educ. Res. **11**, 010114 (2015).

[4] B. Gutmann, G. Gladding, M. Lundsgaard, and T. Stelzer, Mastery-style homework exercises in introductory physics courses: Implementation matters, Phys. Rev. Phys. Educ. Res. **14**, 010128 (2018).

[5] J. H. Block and R. B. Burns, 1: Mastery learning, Review of Research in Education **4**, 3 (1976), https://doi.org/10.3102/0091732X004001003.

[6] B. D. Mikula and A. F. Heckler, Framework and implementation for improving physics essential skills via computer-based practice: Vector math, Phys. Rev. Phys. Educ. Res. **13**, 010122 (2017).

[7] A. F. Heckler and B. D. Mikula, Factors affecting learning of vector math from computer-based practice: Feedback complexity and prior knowledge, Phys. Rev. Phys. Educ. Res. **12**, 010134 (2016).

[8] Z. Chen, C. Chudzicki, D. Palumbo, G. Alexandron, Y.-J. Choi, Q. Zhou, and D. E. Pritchard, Researching for better instructional methods using ab experiments in moocs: results and challenges, Research and Practice in Technology Enhanced Learning **11**, 9 (2016).

[9] Z. Chen, N. Demirci, Y.-J. Choi, and D. E. Pritchard, To draw or not to draw? examining the necessity of problem diagrams using massive open online course experiments, Phys. Rev. Phys. Educ. Res. **13**, 010110 (2017).

[10] M. W. Guthrie and Z. Chen, Adding duration-based quality labels to learning events for improved description of students' online learning behavior, in *Proceedings of the 12th International Conference on Educational Data Mining, EDM 2019, Montréal, Canada, July 2-5, 2019* (2019).

[11] Z. Chen, S. Lee, and G. Garrido, Re-designing the structure of online courses to empower educational data mining, in *Proceedings of the 11th International Conference on Educational Data Mining, EDM 2018, Buffalo, NY, USA, July 15-18, 2018* (2018).

[12] Z. Chen, K. M. Whitcomb, and C. Singh, Measuring the effectiveness of online problem-solving tutorials by multi-level knowledge transfer, in *Physics Education Research Conference 2018*, PER Conference (Washington, DC, 2018).

[13] 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 (Cincinnati, OH, 2017) pp. 84–87.

[14] Z. Berry, I. Turgeon, and F. Yonekura, Obojobo next (2019).

[15] The linear momentum chapter includes one module on the center of mass of a system.

[16] Y. Benjamini and Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, Journal of the Royal statistical society: series B (Methodological) **57**, 289 (1995).

[17] We used ANOVA because there is no non-parametric equivalent that can be easily implemented. These data are also log-normal and thus ANOVA is an appropriate test.

[18] J. D. Bransford and D. L. Schwartz, Chapter 3: Rethinking transfer: A simple proposal with multiple implications, Review of Research in Education **24**, 61 (1999), https://doi.org/10.3102/0091732X024001061.

[19] A. Gavrin, Just-in-time teaching., Published in Metropolitan Universities **17**, 9 (2006).

[20] Although the linear relation for Module 1 is not in the order that we expected and a different statistical test may be more appropriate.

[21] K. R. Koedinger, J. Kim, J. Z. Jia, E. A. McLaughlin, and N. L. Bier, Learning is not a spectator sport: Doing is better than watching for learning from a mooc, in *Proceedings of the Second (2015) ACM Conference on Learning @ Scale*, L@S '15 (ACM, New York, NY, 2015) pp. 111–120.