

# Designing online learning modules to conduct pre- and post-testing at high frequency

Zhongzhou Chen,<sup>1</sup> Geoffrey Garrido,<sup>1</sup> Zachary Berry,<sup>2</sup> Ian Turgeon,<sup>2</sup> and Francisca Yonekura<sup>2</sup>

<sup>1</sup>Department of Physics, University of Central Florida, 4111 Libra Drive, Orlando, FL, 32816

<sup>2</sup>Center for Distributed Learning, University of Central Florida, 12701 Pegasus Drive, Orlando, FL, 32816

We introduce a new type of online instructional design, online learning modules, that effectively allows instructors to conduct pre- and post-testing on the scale of every 20-30 minutes. This paper will focus on estimating students' test-taking effort on the pre-test by analyzing their response time using a multi-component mixture model. In a study involving four online learning modules on mechanical energy, we found that only a small fraction of students display low test-taking effort on the pre-tests. We also show that data from frequent pre- and post-test can provide useful information regarding the instructional effectiveness of the learning materials in each OLM.

## I. INTRODUCTION

Pre and post testing is the single most well established and widely used method for measuring students' learning outcomes in physics education[1,2]. However, most existing pre and post tests are designed to measure students' learning gain over an entire semester, yet most instructors must make a large number of instructional choices on a daily basis [3]. This large frequency gap means that many instructional choices are being made without sufficient data on student learning. In consequence, the instructional design of many physics courses is still being shaped heavily by students' evaluation forms rather than their learning outcomes.

We believe that there are two major hurdles preventing instructors from more frequently administering pre-post assessments. The first is the high cost in time and resources associated with creating, administering and grading an assessment, and the second being the concern that students will find frequent pre-testing disruptive, and as a result, will not have the motivation to take such low-stakes assessments seriously. Research on test-taking effort in low or zero stakes tests found that some students tend to "speed" through tests by answering most of the test items via "rapid guessing"[4,5]. It has also been shown that students' performance on physics conceptual surveys are sensitive to testing conditions [6].

The recent surge in online education and online testing technology provides a potential solution to both hurdles. Most online platforms today allow instructors to draw auto-gradable problems from a large problem bank, greatly reducing the cost of creating and grading assessments. They also allow instructors to be innovative in the format in which the assessment is being administered, so as to boost students' test-taking motivation. More importantly, a well-designed online learning platform can provide rich data on student behavior, allowing researchers to observe and measure students' test-taking effort. [7,8]

In this paper, we introduce a new online instructional design, online learning modules (OLM), that effectively enables pre-post assessment to be conducted on a time scale of every 20-30 minutes. We show that by analyzing students' response time on assessment attempts using a mixture-model method, we can estimate the fraction of students who "speeded" through the pre-test for each OLM. This population turns out to be relatively small in the current study. Meanwhile, the pre and post-test data can provide rich information on the effectiveness of instructional materials in each module.

### A. Design of OLM

Inspired by early research in modularized instructional design and deliberate practice[9,10], each OLM contains instruction, practice problems and assessment (FIG 1 **Error! Reference source not found.**), focused on developing competency in one well-defined "knowledge component"[3]. A knowledge component roughly corresponds to a single physics concept such as kinetic energy, or one aspect of a physics principle, such as conceptual understanding of conservation of mechanical energy. A series of OLMs are combined sequentially to form a learning unit on a certain topic. A student can access the next module in the sequence after he/she passes the assessment of the previous module.

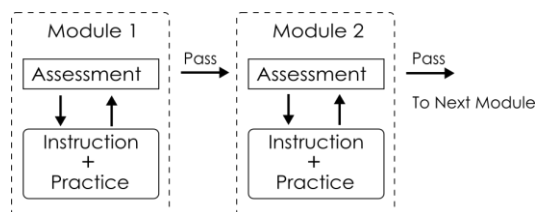


FIG 1: Design of an OLM sequence.

A unique feature of the OLM design is that students are required to attempt the assessment at least once to "unlock"

the instruction and practice problems in each module. After the initial attempt, student can choose to either study the instruction and work on practice problems, or make additional attempts on the assessment. The instruction and practice problems are locked from access during each assessment attempt. For each OLM, the initial assessment attempt serves as a pre-test, while all the attempts afterwards serve as multiple posttests. More detailed discussion of OLM design will be presented in a different paper.

### B. Analyzing response time with mixture models

Students' response times on test items have been shown to correlate well with their test-taking effort on not for credit tests [7,8]. More specifically, students who speed through a test by answering items via rapid-guessing would spend significantly less time on the test, showing up as a peak close to zero in the response time distribution. The sizes of different student groups with different test-taking behavior can be estimated by fitting the response time distribution to a mixture-model with  $m$  components:

$$g(t) = \sum_{i=1}^m \lambda_i \phi_i(t)$$

where  $g(t)$  is the probability density at response time  $t$ ,  $\lambda_i$  are the relative weights of the components which sum to unity, and  $\phi_i(t)$  is the density function of component  $i$ . Each component would ideally correspond to a group with a distinct test-taking behavior, with a maximum of  $m$  groups thought to exist in the population. A student with response time  $\tau$  is assigned to group  $j$  which corresponds to the component that has the maximum weighted probability density among all components:  $\lambda_j \phi_j(\tau) = \max[\lambda_i \phi_i(\tau)]$ .

Since the assessment of each OLM only contains about 2-3 problems, in the current study we use the total response time on one assessment attempt, referred to as the attempt response time (ART), as a proxy for students' test-taking effort.

## II. METHODS:

### A. Creation and Implementation of OLM sequence

For the current study we created an OLM sequence consisting of four modules on the topic of conservation of mechanical energy(CME), including: definition of kinetic energy(KE), definition of gravitational and elastic potential energy (PE), conceptual understanding of CME (CU) and problem solving using CME (PS). The learning modules are implemented in an online learning platform, Obojobo, developed by the Center for Distributed Learning at University of Central Florida [11].

The assessment component of each module contains three sets of 2-3 isomorphic assessment problems, inspired by or directly taken from either the Energy and Momentum Conceptual Survey [12], or an exam review instrument developed by the PER group at University of Illinois [13]. Students are presented with one of the three sets on each assessment attempt in fixed order. After each attempt, students are informed of the correctness of their answer to each problem, but not the correct answer itself. A student passes an assessment when he/she can correctly answer all questions on a single attempt. Since students cannot access subsequent modules before passing the assessment, they are allowed 20 attempts on each assessment in the current experiment.

The instructional component consists of instructional text and images, interleaved with practice problems which provide students with wrong-answer feedback and problem solution after each attempt.

### B. Experiment Setup

Student subjects were recruited from a calculus based introductory mechanics course of a large south-eastern public university. Subjects were given access to the OLM sequence as an exam review tool one week before the midterm exam that covered CME. No course credit was assigned for completing the OLM sequence.

### C. Data Collection and Analysis

Click-stream data from subjects were collected from the Obojobo platform after the experiment and analyzed using the software suite R [14]. ART is defined as the time between the start and end of an assessment attempt, marked by two distinct mouse-click events. Analysis of ART on initial attempt using mixture-model is conducted with the R package mixtools [15].

## III. RESULTS

A total of 77 students registered for the study, of which 75 launched the assessment of the first module. We first attempted to model the distribution of ART of initial assessment attempt using a two component log-normal distribution following the method outlined in [7].

However, for three of the four cases this method resulted in an unexpected best fitting model where one of the two components accounted for both the "speeded" group on the left and the "slow" group on the right, which contains students that took an exceptionally long time to complete the assessment (the red curve in FIG 2). This result prevented us from estimating the size and mean ART of the "speeded" group alone, and there is no reason to believe that "speeded"

and “slow” groups have similar test-taking effort. Adding a third component to account for the “slow” group failed to resolve the problem and resulted in similar outcomes.

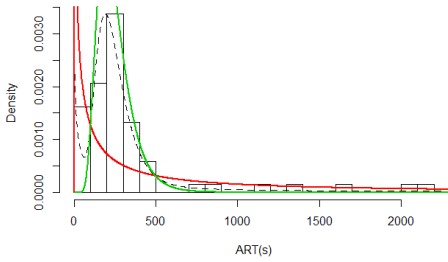


FIG 2: Example of a 2 component log normal fit of student ART data on initial attempt.

To properly separate the “speeded” and “slow” groups, we adopted a 3-component normal distribution mixture model of the form:

$$\phi_i(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \times \exp\left\{-\frac{(t - \mu)^2}{2\sigma^2}\right\}$$

The three components were intended to capture the “speeded”, “normal” and “slow” test taker groups respectively. To further exclude the impact of exceptionally long ART data which is highly non-normal, the longest 10% of ART for each assessment is excluded from the analysis. This new model is likely to improve the validity of the

TABLE 1: Resulting parameters of the 3-component normal distribution mixture model fit. Coding for the three components: 1 = speeded, 2 = normal, 3 = slow

Module	KE			PE			CU			PS		
Loglikelihood	-388.80			-268.75			-253.09			-243.61		
Components	1	2	3	1	2	3	1	2	3	1	2	3
$\mu$	<b>62.64</b>	243.67	802.00	<b>32.34</b>	162.89	371.12	<b>17.58</b>	161.88	371.58	<b>32.55</b>	474.83	1299.42
$\sigma$	<b>38.75</b>	95.00	19.00	<b>17.62</b>	63.96	27.15	<b>6.34</b>	56.96	90.23	<b>17.93</b>	296.06	54.42
$\lambda$	<b>0.13</b>	0.84	0.03	<b>0.16</b>	0.78	0.06	<b>0.09</b>	0.52	0.39	<b>0.11</b>	0.84	0.05
$n$	<b>8</b>	53	2	<b>9</b>	34	3	<b>4</b>	23	14	<b>4</b>	28	2

#### IV. DISCUSSION

We have shown that the size of the “speeded” test-taker population on frequent OLM pre-tests can be estimated using a mixture-model analysis method, and that this population is relatively small for the OLMs involved in the current study.

Several design features of OLMs might have contributed to the small number of “speeded” test-takers. First of all, the fact that students can directly proceed to the next module without having to go through the rest of the current module may have provided the internal motivation for high test-taking effort. Secondly, the short length of each pre-test may have prevented the decline of test-taking effort, as it has been shown to decrease with the length of the test in some cases [5]. Finally, the multiple attempt design of OLMs might have created a game-like environment, in which the assessment is

method as it improves the fitting of the peak close to zero in the data.

The resulting three-component models are plotted in FIG 3 and the parameters displayed in TABLE 1. The estimated number of “speeded” test-takers is less than 20% of the population for module PE, and < 15% for the other three modules.

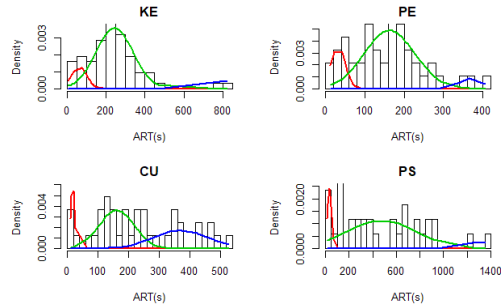


FIG 3: 3-component log-normal fit of ART on initial assessment attempt.

In FIG 4, we plot the number of students who attempted each module, grouped by the number of attempts taken to pass the module. It is worth noting that the problem sets repeat every three attempts, meaning that students who passed the assessment on > 3 attempts benefited from previous attempts.

being viewed as a challenge similar to the level-boss in a video game.

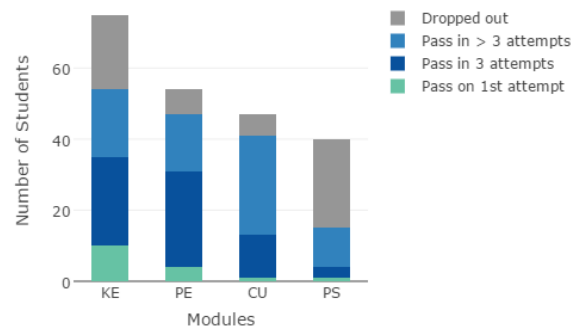


FIG 4: Number of students passing each module grouped by number of attempts.

However, we must also note that the experiment is conducted as a review unit after classroom instruction on the content, and that participation is voluntary. Having been exposed to the content before could have boosted the confidence of students on the pre-test, although the highest first attempt passing rate among all four modules is less than 20%. In addition, unmotivated students might simply have dropped out of the study altogether without making an initial attempt. It is possible that the “speeded” test-taker population will increase if the OLMs are assigned for credit in a course.

Nonetheless, the OLMs enabled us to closely monitor students’ test-taking effort, and study how it changes with different test administration conditions. To the best of our knowledge, this is the first study to measure students’ test-taking effort on a physics pre-test using students’ response times.

More importantly, the multi-attempt pre- and post-assessments of OLMs not only measure the level of mastery for each knowledge component, but also measure the effectiveness of the instructional materials in each OLM. As shown in FIG 4, modules KE and PE have higher effectiveness, as reflected by the number of students who passed the modules in less than 3 attempts (dark blue). In contrast, few students passed module PS after studying the module, indicating that its instructional materials require significant improvement. The data also shows that more students tend to drop out on the first and last modules of the sequence. This might be caused by lack of incentive for students to complete the modules, although the real reason will be an interesting topic for future research

By providing rich data on students’ learning from online instructional materials on the scale of 20-30 minutes, OLMs can serve as a valuable complement to the traditional pre and

post concept tests, which measure learning gains from classroom instruction over much longer timescale.

## V. LIMITATIONS AND FUTURE DEVELOPMENTS

We noticed a few limitations in the current implementation of OLMs which can be improved in the future. For one, to ensure that every student can access all four modules, we had to give 20 attempts on every assessment to ensure that every student can pass the modules. In future implementations, we will give a smaller number of attempts, and allow students to access subsequent modules once all the attempts in the previous module were used up. A second limitation is that the current mixture model analysis relies on a pre-determined number of components. In future implementations with a larger sample size, an iterative bootstrapping likelihood ratio test can be used to determine the optimum number of components for a given distribution based on fitting indices such as AIC and BIC [15]. Furthermore, in future studies the validity of the data analysis method can be further examined by surveying students about their test taking effort, as is done in [16].

Finally, it will be an interesting future direction to study how students’ test taking behavior change when OLMs are assigned for course credit, as well as how negative effects of over-testing can be avoided if we wish to use OLMs more extensively as a tool for formative assessment.

## ACKNOWLEDGEMENTS

The authors would like to thank University of Central Florida and the Center for Distributed Learning for supporting the current research, Dr. Sunbok Lee for helpful discussion on mixture-model techniques.

- 
- [1] Hake R R 1998 *Am. J. Phys.* **66** 64–74
  - [2] Meltzer D E 2002 *Am. J. Phys.* **70** 1259–68
  - [3] Koedinger K R, Booth J L and Klahr D 2013 *Science (80-. )*. **342** 935–7
  - [4] Barry C L, Horst S J, Finney S J, Brown A R and Kopp J P 2010 *Int. J. Test.* **10** 342–63
  - [5] Wise S L, Pastor D A and Kong X J 2009 *Appl. Meas. Educ.* **22** 185–205
  - [6] Ding L, Reay N W, Lee A and Bao L 2008 *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **4** 2–7
  - [7] Schnipke D L and Scrams D J 1999 *J. Educ. Meas.* **34** 213–32
  - [8] Wise S L and Kong X 2005 *Appl. Meas. Educ.* **18** 163–83
  - [9] Goldschmid B and Goldschmid M L 1973 *High. Educ.* **2** 15–32
  - [10] Ericsson K 2006 *The Cambridge handbook of expertise and expert performance* ed K A Ericsson, N Charness, P J Feltovich and R R Hoffman (America) pp 683–703
  - [11] Bishop C, Yonekura F and Moskal P 2013 *Proceedings of the Association of College and Research Libraries (ACRL) Annual Conference* (Indianapolis, Indiana) pp 466–71
  - [12] Singh C and Rosengrant D 2003 *Am. J. Phys.* **71** 607
  - [13] Fakcharoenphol W, Potter E and Stelzer T 2011 *Phys. Rev. Spec. Top. - Phys. Educ. Res.* **7** 1–7
  - [14] R Core Team 2017 *Doc. Free. available internet* <http://www.r-project.org>
  - [15] Benaglia T, Chauveau D, Hunter D and Young D 2009 *J. Stat. Softw.* **32** 1–29
  - [16] Follette K, McCarthy D, Dokter E, Buxner S and Prather E 2015 *Numeracy* **8**