

When Students Can Choose Easy, Medium, Or Hard Homework Problems

Raluca E. Teodorescu, Daniel T. Seaton, Caroline N. Cardamone, Saif Rayyan, Jonathan E. Abbott, Analia Barrantes, Andrew Pawl*, and David E. Pritchard

Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139.

**Department of Chemistry and Engineering Physics, University of Wisconsin-Platteville, Platteville, WI 53818.*

Abstract. We investigate student-chosen, multi-level homework in our Integrated Learning Environment for Mechanics [1] built using the LON-CAPA [2] open-source learning system. Multi-level refers to problems categorized as *easy*, *medium*, and *hard*. Problem levels were determined a priori based on the knowledge needed to solve them [3]. We analyze these problems using three measures: time-per-problem, LON-CAPA difficulty, and item difficulty measured by item response theory. Our analysis of student behavior in this environment suggests that time-per-problem is strongly dependent on problem category, unlike either score-based measures. We also found trends in student choice of problems, overall effort, and efficiency across the student population. Allowing students choice in problem solving seems to improve their motivation; 70% of students worked additional problems for which no credit was given.

Keywords: Introductory Physics, Problem Solving, Online Homework

PACS: 01.40.Fk, 01.40.Ha

INTRODUCTION

Problem classifications such as *easy*, *medium*, and *hard* are widely used in physics textbooks and curricular materials. Such classifications are typically based on instructors' opinion with little research-based justification. In our Integrated Learning Environment for Mechanics (ILEM), we are investigating quantitative measures that justify such classifications, as well as students' behavior when allowed to choose from a collection of pre-categorized problems. This research will be used to enhance our categorization and the pedagogical effectiveness of ILEM.

In this article, we investigate problem characteristics and student performance using multiple assessment techniques. We aim to measure properties of our categories and to understand how students use the system.

BACKGROUND

We implemented student-chosen multi-level homework during a reformed off-semester introductory mechanics course in Spring 2011 (8.011) for students who had failed or not completed introductory mechanics (8.01). Student-chosen homework was offered only in the first part of the course, targeting basic knowledge and solving single-concept problems; the latter part of the course emphasized developing strategic knowledge necessary to identify physical principles in multi-concept problems. Student-chosen homework problems were or-

ganized in *easy*, *medium*, and *hard* categories using the Taxonomy of Introductory Physics Problems [3] that categorizes problems based on cognitive processes and knowledge required. Our problems involved similar cognitive processes, but different declarative and procedural knowledge (Table 1 has characteristics and examples).

Each homework assignment in ILEM featured an average of six problems from each category weighted by points: *easy*=1pt, *medium*=2pts, and *hard*=3pts. Students were required to accumulate at least 15 pts with no restrictions regarding the types or order of problems solved. Students were allowed 7 attempts to solve each problem and it was always possible to accumulate 15 points while working within only two of the categories. A total of five assignments were given along with a wiki-based online text to introduce the relevant material. The ILEM homework was due before students came to class and was meant to establish their basic knowledge of the physical concepts and their understanding of the introductory text. After class, students had to complete Mastering Physics [4] problems and written problem sets. The course was taught in a SCALE-UP [5] format with ~18 students per section using our Modeling Applied to Problem Solving pedagogy [6, 7]. Classes emphasized students solving problems in pairs at the board.

DATA ANALYSIS AND RESULTS

Our homework has two unique features: problems are classified according to cognitive processes and knowl-

TABLE 1. The cognitive processes and the knowledge (both declarative and procedural) targeted by *easy*, *medium*, and *hard* problems. For more detailed definitions of the cognitive processes and knowledge types see ref. [3].

Difficulty	Cognitive Processes	Declarative Knowledge	Mental Procedures	Examples
Easy	recall, execution, representation, ranking, analyzing errors	definitions, vocabulary terms, basic facts, simple time sequences	single rules and basic algorithms	evaluate definitions, identify appropriate systems, perform simple calculations, match basic graphs with verbal descriptions, match arrows with relevant forces
Medium	same as Easy + integration	complex facts and time sequences	complex algorithms and tactics	choose an appropriate problem solving strategy, perform two-step calculations, compare physical quantities and outcomes, over-informed scenarios
Hard	same as Easy + integration	more complex facts and time sequences, principles and generalizations	complex procedures	evaluate solutions, match complex diagrams with verbal descriptions, match problems with their strategies, perform multi-step and limiting case calculations, multiple object scenarios

edge required and students choose which problems to solve. These features raise the following questions:

- How do our qualitative problem categories compare with performance-based measures of difficulty?
- How much time do students spend on the different types of problems and with what efficiency?
- What seems to guide students' choice as to which and how many problems to do?

Three assessment techniques have been used to address these questions: LON-CAPA difficulty $Diff_{LC}$, item difficulty using item response theory ($Diff_{IRT}$), and time measures describing both problems and students. All submissions in LON-CAPA are logged with a time stamp, a score (right or wrong), and the number of attempts, allowing performance-based and time-based measurements for problems and students. A total of $N = 63$ students interacted with $M = 99$ questions embedded in our multi-level homework in ILEM.

LON-CAPA estimates problem difficulty using methods similar to classical test theory, defining the difficulty as the fraction of submissions that are incorrect, $Diff_{LC} = \text{Incorrect}/\text{Total}$ ($Diff_{LC} = 0$ if all submissions are correct and $Diff_{LC} = 1$ if all submissions are incorrect). $Diff_{LC}$ is based only on those students who attempted a given problem and so may be biased towards students with either high or low average skill.

IRT allows for a skill-based assessment of problem difficulty, where the probability that a student of a given skill level would answer each problem correctly is fit using logistic models. For these calculations, a matrix of student responses for each problem is dichotomized into correct/incorrect (or no response). In the IRT analysis presented here, we consider only the first submission to a problem; little variation was found between levels

when applying IRT to the seventh submission. We provide results from a simple Rasch model [8] using the open source package in R known as *ltm*. Since our course was relatively small, problem discriminations could not be performed; details regarding IRT and our implementation can be found in ref. [9].

Time-based measures are calculated using time stamps associated with the first view and subsequent submissions of a problem. We define the time-per-problem ($\tau^{problem}$) as the average time students spend on a particular problem (times for multiple attempts are accumulated), irrespective of whether the problem was answered correctly. We define the time-per-submission ($\tau^{submission}$) as the average time of individual answer submissions on a certain problem, again, irrespective of whether the submission was correct. In all our time-based calculations, we exclude submission times longer than 30 minutes to account for idle students; this cutoff removes $\approx 4\%$ of the total number of submissions.

Distinguishing Easy, Medium, And Hard

We first examine the differences among our *easy*, *medium*, and *hard* categories of problems with respect to student performance and time spent. In Fig. 1, two types of problem difficulty ($Diff_{LC}$ and $Diff_{IRT}$), time-per-problem ($\tau^{problem}$), and time-per-submission ($\tau^{submission}$) are plotted for each *easy*, *medium*, and *hard* problem category. Within each measure, our primary interest is the relative position of the plotted quantities.

Somewhat to our surprise, both measures of problem difficulty show little difference across the categories, especially between medium and hard. The $Diff_{LC}$ (top plot in Fig. 1) has small variation between *easy*, *medium*, and *hard* categories: $\sim 58\%$ of all answers submitted to

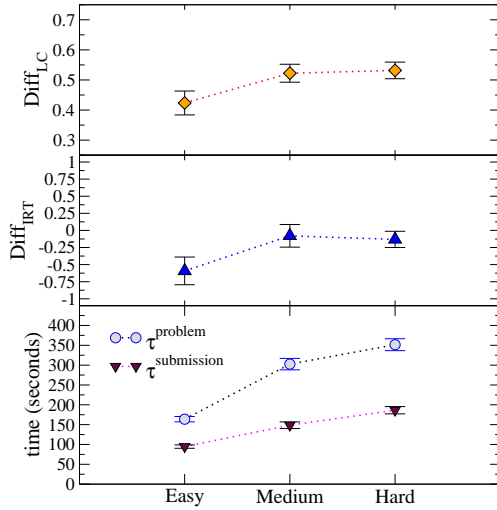


FIGURE 1. The LON-CAPA difficulty ($Diff_{LC}$), IRT difficulty ($Diff_{IRT}$), time-per-problem ($\tau^{problem}$), and the time-per-submission ($\tau^{submission}$) for *easy*, *medium*, and *hard* problems. Error bars in all quantities reflect standard error of the mean and connecting lines are to guide the eye.

easy problems are correct vs. $\sim 48\%$ for both medium and hard problems. It is important to consider student-choice when analyzing $Diff_{LC}$ because hard problems were typically done by the better students, leading to values that may differ from the case where all students attempted a given problem. IRT analysis confirmed the $Diff_{LC}$ measures by giving similar results: medium and hard problems have $Diff_{IRT} \sim -0.1$ whereas the easy problems have $Diff_{IRT} \sim -0.6$. On first try, average students answer medium and hard problems correctly $\sim 52\%$ of the time vs. $\sim 65\%$ easy problems. IRT takes into account student skill level when determining problem difficulty, so our medium and hard problems are roughly equal in difficulty as judged by the probability of students answering them correctly.

In contrast, the bottom plot in Fig. 1 shows significant differences in both the time-per-problem ($\tau^{problem}$) and the time-per-submission ($\tau^{submission}$) across the three problem categories. These measures show that the more complex the problem's procedural and declarative knowledge (as displayed in Table 1), the more time needed for a student to solve the problem and the more time needed for each submission.

Student Choice Behavior

We now look at student behavior in selecting problems. Of the total number of problems attempted by our class, 45% were easy, 35% were medium, and 20% were

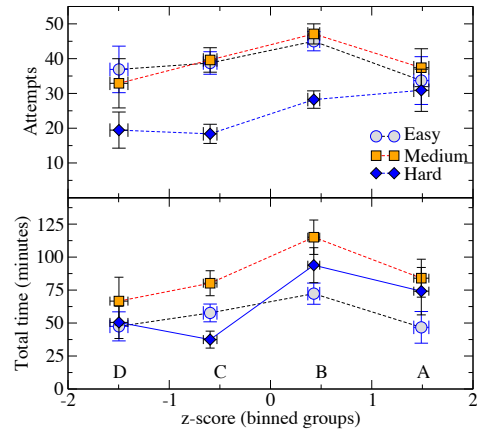


FIGURE 2. Attempts per student and total time per student for the three problem categories, averaged within the four z-score groups. Horizontal error bars represent the standard error of the z-score within each bin and vertical error bars represent the standard errors of the quantities plotted on the y-axes.

hard, yielding 26%, 40%, and 34% of the total points earned respectively. In order to better understand student behavior, we binned students in categories based on exam performance (weekly quizzes and final), and analyzed additional parameters related to their effort, strategy, efficiency, and motivation.

We define four groups of students, expressing their total exams grade as z-scores (standard deviations from the mean, e.g. a student with a z-score of +1.5 has grade that is 1.5 standard deviations above the average of the class). These four groups, labeled A (the best performing), B, C, and D (the worst performing) have the following ranges:

- A ($1.0 < z\text{-score} \leq 2.0$, 11 students)
- B ($0.0 < z\text{-score} \leq 1.0$, 16 students)
- C ($-1.0 < z\text{-score} \leq 0.0$, 26 students)
- D ($-2.0 < z\text{-score} \leq -1.0$, 10 students)

Fig. 2 provides measurements of effort for students in each of the four groups: their total number of attempts (right or wrong) for each problem category and the total time spent working problems in each category. The most striking feature in Fig. 2 is that both the total number of attempts and the total time spent working problems in each category rise from group D to group B, peaking with group B. Group A behaves differently, where a larger fraction of attempts and time is spent on hard problems, but total time and total attempts are down with respect to other groups. Group A appears to be able to exceed the required number of points with less total effort, confounding the general belief that A students work harder. This perceived low effort may reflect that LON-CAPA assignments were straightforward and prepara-

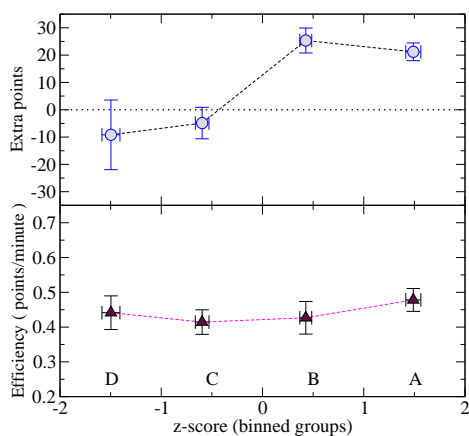


FIGURE 3. Extra points (total points earned minus the 75 required points) and efficiency (total points earned per minute), averaged within the four z-score groups. Horizontal error bars represent standard error of the z-score within each bin and vertical error bars represent standard errors of the quantities plotted on the y-axes.

tory, designed to bring students up to a common foundation for the next class. (Mastering Physics homework [4] and written assignments were more challenging.)

We now investigate measures more indicative of student strategy and efficiency. In Fig. 3, we plot the extra points earned after fulfilling the assignment requirements (top) and student efficiency defined as the total points earned per minute (bottom), both averaged within the four z-score groups. The top plot shows that B and A students earn many extra points through all assignments, while C students tend to earn only the required points and D students often fall short of the requirements.

Surprisingly, the points per minute measure is equal within error across all groups. Our efficiency measure suggests that choice allows students to earn a reasonable amount of points with problems that match their ability. It seems groups C and D may simply elect not to invest more time when their success rate drops, even when sufficient points have not been accumulated.

CONCLUSIONS

We categorized problems based on the types of declarative and procedural knowledge involved as *easy*, *medium*, and *hard* (all were single-concept problems involving similar cognitive processes). We offered students choice in the type of problems they solved and discovered that time-per-problem and time-per-submission are strongly dependent on the *easy*, *medium*, and *hard* categories, whereas LON-CAPA difficulty and IRT item difficulty are less dependent. Furthermore, we found that

allowing students choice in problem solving leads to the following patterns of behavior:

- D, C, and B students spent progressively more total time solving problems whereas A students satisfied requirements with effort comparable to D students.
- 70% of students earned extra points on assignments (consistent with previous findings on flexible versus traditional homework [10]).
- All students (regardless of z-score) have comparable efficiencies in solving problems.

In the future, it will be interesting to consider the effect of the contexts [11], the students' self-awareness, epistemological beliefs [12], and familiarity (experience) [13] with their problem-solving behavior.

ACKNOWLEDGMENTS

We are grateful to G. Kortemeyer for help in interpreting LON-CAPA activity logs. This project was supported by NSF grants PHY-0757931 and DUE-1044294.

REFERENCES

1. R. E. Teodorescu, A. Pawl, S. Rayyan, A. Barrantes, and D. E. Pritchard, in *Physics Education Research Conference 2010* (Portland, Oregon, 2010), vol. 1289 of *PER Conference Invited Paper*, pp. 321–324.
2. www.loncapa.org.
3. R. E. Teodorescu, Ph.D. thesis, George Washington University, Washington D.C., D.C. (2009).
4. www.masteringphysics.com.
5. R. J. Beichner, J. M. Saul, D. S. Abbott, J. J. Morse, D. L. Deardorff, R. J. Allain, S. W. Bonham, M. H. Dancy, and J. S. Risley, in *Research-Based Reform of University Physics* (2007), vol. 1.
6. A. Pawl, A. Barrantes, and D. E. Pritchard, in *Physics Education Research Conference 2009* (Ann Arbor, Michigan, 2009), vol. 1179 of *PER Conference Invited Paper*, pp. 51–54.
7. S. Rayyan, A. Pawl, A. Barrantes, R. Teodorescu, and D. E. Pritchard, in *Physics Education Research Conference 2010* (Portland, Oregon, 2010), vol. 1289 of *PER Conference*, pp. 273–276.
8. R. de Ayala, *The Theory and Practice of Item Response Theory* (The Guilford Press, 2008), 1st ed., ISBN 1593858698.
9. C. Cardamone, J. Abbott, A. Barrantes, A. Pawl, S. Rayyan, D. T. Seaton, R. Teodorescu, and D. Pritchard (2011), submitted to Proceedings of the 2011 Physics Education Research Conference.
10. L. Bao, S. R. Stonebraker, and H. Sadaghiani, *Am J Phys* **76**, 878 (2008), ISSN 00029505.
11. M. S. Sabella, Ph.D. thesis, University of Maryland, College Park, MD (1999).
12. J. C. Stockton, Ph.D. thesis, Kennesaw State University, Kennesaw, GA (2010).
13. R. Marzano and J. Kendall, *The new taxonomy of educational objectives* (Corwin Press, 2007), ISBN 9781412936293, page 11.