Beta-Test Data On An Assessment Of Textbook Problem Solving Ability: An Argument For Right/Wrong Grading?

Karen Cummings* and Jeffrey D. Marx#

*Southern Connecticut State University, 501 Crescent St., New Haven, CT 06515
#McDaniel College, Department of Physics, Westminster, MD 21158

Abstract. We have developed an assessment of students' ability to solve standard textbook style problems and are currently engaged in the validation and revision process. The assessment covers the topics of force and motion, conservation of momentum and conservation of energy at a level consistent with most calculus-based, introductory physics courses. This tool is discussed in more detail in an accompanying paper by Marx and Cummings. [1] Here we present preliminary beta-test data collected at four schools during the 2009/2010 academic year. Data include both pre- and post-instruction results for introductory physics courses as well as results for physics majors in later years. In addition, we present evidence that right/wrong grading may well be a perfectly acceptable grading procedure for a course-level assessment of this type.

Keywords: Problem solving in physics, assessment, introductory physics, right/wrong grading.
PACS: 01.40.Fk

INTRODUCTION

Marx and Cummings have developed an assessment of students' ability to solve standard textbook style problems and are currently engaged in beta-testing the instrument in preparation for revision and eventually formal statistical validation. The assessment covers the topics of force and motion, conservation of momentum and conservation of energy at a level consistent with most calculus-based, introductory physics courses. This tool is discussed in more detail in an accompanying paper by Marx and Cummings. However, one detail important for this paper is that the exam is written so that questions within a given topical cluster (e.g. conservation of momentum) should increase in difficulty as the exam progresses.

Here we present preliminary data collected from four universities during the 2009/2010 academic year. Data include both pre- and post-instruction results for introductory physics courses as well as results for physics majors in later years. In all cases, the data were collected using an open-ended format of the current version of the assessment. Hence, we present results based both on written student work and final answers alone in the discussions below.

The goal of this paper is to provide some insight into three important issues. 1) Do preliminary data indicate, in a general way, that the exam has potential as a measure of students' ability to solve problems? 2) Do the clusters of questions (force and motion (F), conservation of momentum (p), and conservation of energy (E)) show progressing difficulty and an appropriate range of difficulty levels? 3) What do the open ended responses collected indicate about possible future directions for large population, objective grading of this diagnostic tool?

Due to space limitations, not all data collect are presented here. Only examples of data directly relevant to the discussion are shown.

BACKGROUND ON TESTING SITES

As mentioned previously, data were collected from four universities. These schools span a reasonable range of size and selectivity. Their general characteristics are shown in Table 1. A total of approximately 275 individual students took the exam at least once. Many took the exam both pre-instruction and post-instruction in a calculus-based introductory physics course. Table 2 shows information regarding testing at the four schools.

Two of the testing sites were the home institutions of the authors. One other school involved has an active PER group within the physics department. This information is summarized in Table 3.

TABLE 1. Information on schools involved in the Beta-test. School codes are used to protect privacy. We will disclose school names in private communications.

School Code	Public/ Private	Selectivity
School A	Private R1	Med-High
School B	Public R1	Med
School C	Public Comprehen-sive	Low
School D	Private Liberal Arts	Med

TABLE 2. Courses in Which Data Were Collected.

School Code	Calc-Based Physics I?	Other Courses?
School A	PRE/POST	NONE
School B	POST ONLY	NONE
School C	PRE/POST	2nd year physics majors (Modern Physics) Sr physics majors (Research Methods Course)
School D	PRE/POST	NONE

TABLE 3. Information on faculty PER involvement.

School Code	Author Taught Students in Data set?	PER Activity in the Department?
School A	NO	YES
School B	NO	NO
School C	YES	YES
School D	YES	YES

STUDENT PREFORMANCE ON THE ASSESSMENT BETA-TEST

One issue that is of primary importance to us is whether the exam data are generally consistent with what one might expect. Do students at more selective schools pretest higher? Do students' scores improve even slightly as a result of instruction? Do upper level physics majors compare with students just completing Physics I in a way that we might understand? Figures 1-3 show relevant examples of data on student performance on the exam.

School A is a much more selective institution than is School C. Figures 1 and 2 show pre-test scores are much lower at School C, consistent with what one would expect to measure with a valid instrument. However, students at School C have one of the authors for an instructor and research-based methods of teaching problem solving are used (Cooperative Group Problem Solving for example). School C has a higher post-test average which is, again, what one might predict.

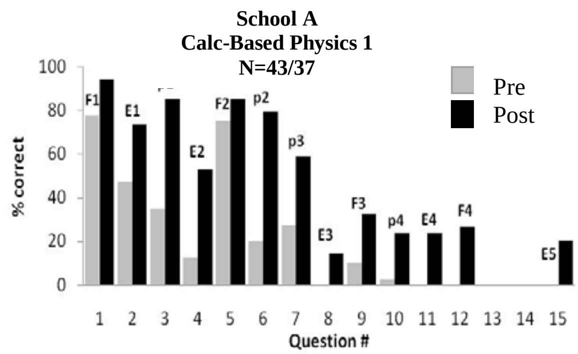

FIGURE 1. % correct on each of the 15 questions for a selective private R1 school. An answer was scored as correct if and only if the numerical answer was correct. Students were told not to answer #13 and #14 for instructional reasons. Force problems are noted with an F, conservation of momentum problems with a p and conservation of energy problems with an E. Problems increase in difficulty from level 1 (F1, E1, p1) to level 5 (F5, E5, p5).

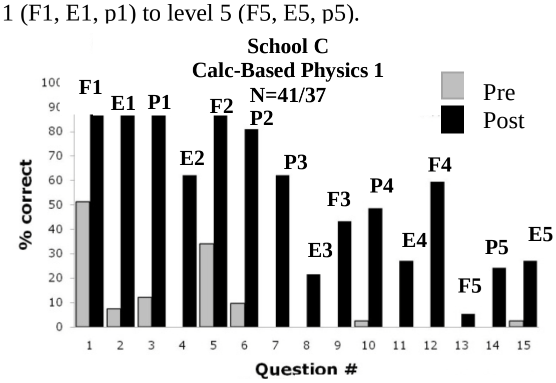

FIGURE 2. Percent correct for a much less selective public institution with one of the authors as instructor. Again, answers are scored as correct if and only if the numerical answer is correct.

Figure 3 shows how students at various points in a physics program (first year, second year, forth year) perform on the exam. Again, we consider these results reasonable given our assumption that student performance should improve with increasing years in the major, but this effect would be confounded by some lack of retention and interference from new, related knowledge.

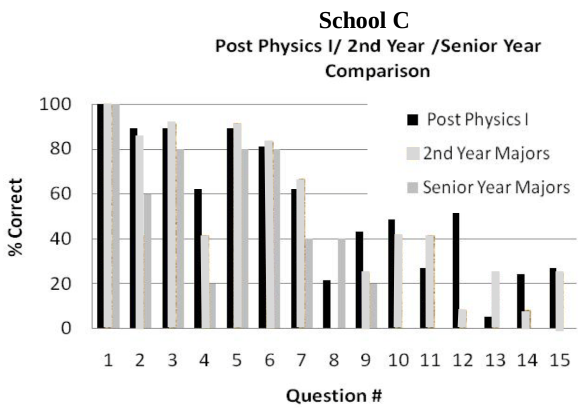

FIGURE 3. A comparison between student performance at the end of physics I, the beginning of the second year and senior year of the physics major.

CONTENT CLUSTERS: EVIDENCE OF PROGRESSING ITEM DIFFICULTY?

A second issue we discuss here is whether the preliminary data indicate that the questions get progressively more difficult within in each cluster. This is an important part of the test design (discussed in more detail in the companion paper). Overall percent correct for all physics I students following instruction is shown in Fig. 4-broken out by cluster.

We are pleased with the results for the conservation of momentum cluster. However, we feel that the force cluster needs work in several regards. It seems the first question (or two) is/are too easy while # 9 (F3) and # 13 (F5) appear to be too difficult.

Unfortunately, the conservation of energy cluster may also need to be revised. Many of these questions can be solved using alternative methods and students often do this as evidenced by their written responses. While we are not philosophically opposed to students using alternative methods, this will mean that exam factors will not load cleanly on the original three stated domains. This is an issue we need to consider further. One question, #4 (E2), can be solved using an incorrect assumption yet yielding a correct answer, so has to be cut or reworked.

RIGHT/WRONG GRADING: A VIABLE ALTERNATIVE TO MULTIPLE CHOICE?

Our goal for this assessment has always been to provide a tool which is reasonably easy to use, even in large enrollment courses. Hence, our hope was to use the open-ended responses we gathered in this beta-test to generate distracters for a multiple choice exam. Unfortunately, we now believe that the multiple choice option is unlikely. The range of incorrect answers we observed would require too large of a number of distracters. Fortunately, a new alternative, right/wrong grading, is looking very hopeful. This optimism is based on Fig. 5 below which shows scores on the exam when graded by instructors considering all written work and assigning partial credit as they would normally do versus scores when final answers are simply evaluated as right or wrong. The correlation between these two scoring techniques is very strong so, given we have few choices, we intend to move forward with a right/wrong grading approach

However, careful evaluation of Fig. 5 shows that right/wrong grading would be an unacceptable approach to evaluating individual student performance. That is, right/wrong grading of this exam would be recommended only if the exam is used as intended-to evaluate courses.

SUMMARY

We are generally optimistic about the future of this instrument following our collection of preliminary data. Hopeful signs include the fact that scores improve following instruction but not to the point where the score curves are truncated at the high end due many 100% correct scores. However, we would like to give this assessment in a few more selective universities before finalizing that determination.

Scores among the student groups (School A vs. C for example) are consistent with other measures of student ability like SAT scores and previous physics experience. Scores among physics majors in later courses are much higher than pre-instruction scores for Physics I students but lower than for Physics I students post-instruction. This is consistent with the decay curve typically seen in retention investigations.

Face validity among instructors who have worked with the exam seems reasonable. But this issue, along with data on the time needed for instructors to complete the exam, is an area in which we desperately need colleagues to assist us by volunteering to take the exam themselves.

Multiple-choice grading of the exam is beginning to looking impossible-even to us- but we believe we have an acceptable alternative in right/wrong grading (only acceptable for course level evaluation, not acceptable for the evaluation of individual students). This opens the door to creative use of scan forms or web-based homework systems for exam administration

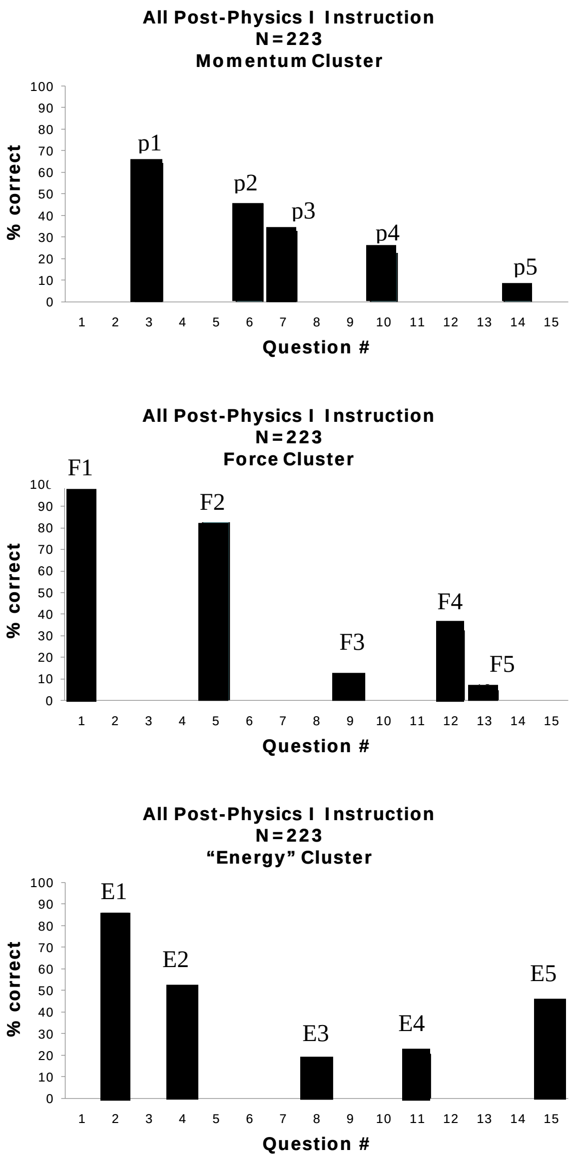

FIGURE4 A-C. % correct on all questions within a given cluster for all post-instruction physics I students for each of the three clusters. We would like these data to indicate progressing question difficulty level. They seem to do so in the momentum cluster.

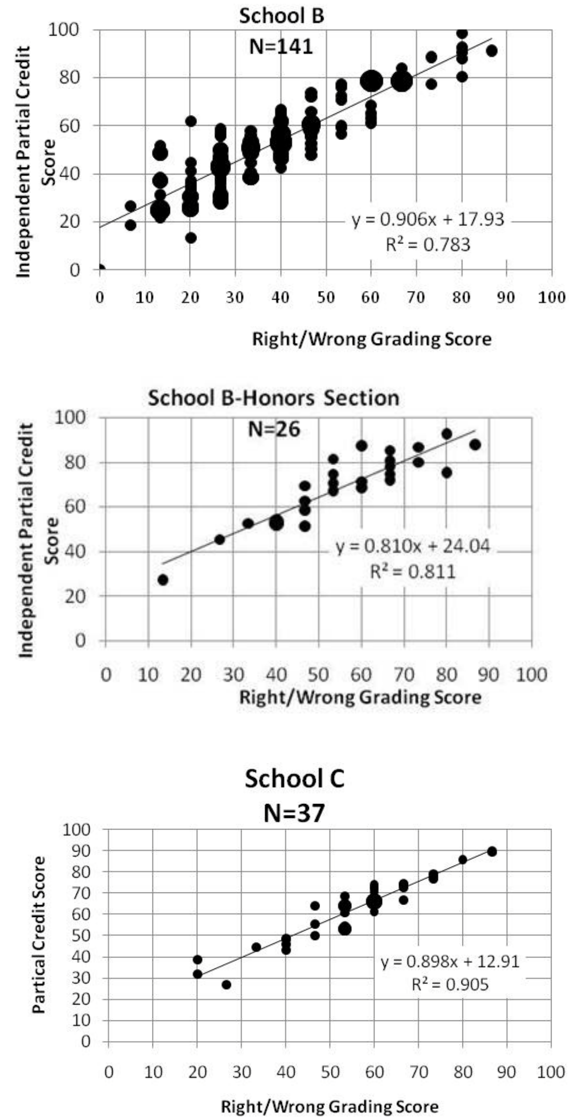

FIGURE 5. A comparison between scores on the exam when all written work is evaluated and graded by the instructor and scores when final answers are simply marked right or wrong. Correlations are strong in all groups for which these data were available.

REFERENCES AND NOTES

1. Marx and Cummings, Physics Education Research Conference Proceedings, 2010, In Press.
2. *Contact e-mail information*
Jeff Marx: jmarx@mcdaniel.edu
Karen Cummings: cummingsk2@southernct.edu