Development of a Survey Instrument to Gauge Students' Problem-Solving Abilities

Jeffrey Marx and Karen Cummings

McDaniel College, Department of Physics, Westminster, MD 21158
Southern Connecticut State University, Department of Physics, New Haven, CT 06515

Abstract. In this paper we discuss the early stages of development of a survey instrument to assess students' problem-solving abilities in a first-term, undergraduate, calculus-based physics course. Specifically, we present our motivation for the development of such a survey, details of a preliminary version of the survey, and some sample items.

Keywords: Problem-solving, assessment, introductory physics.
PACS: 01.40.Fk, 01.40.G-

INTRODUCTION

An oft-stated and embraced outcome of many introductory physics courses, apart from the obvious content expectations, relates to improving students' problem-solving abilities. The aim of this paper is to describe the motivation for, provide an overview of, and display examples from the design and development of a survey instrument intended to gauge students' problem-solving ability in a first-term, undergraduate, calculus-based physics course.

Physics education researchers have allotted significant resources investigating, at various levels, problem-solving ability in students and developing curricular materials intended to improve such abilities.[1-7] Different research projects related to problem-solving have adopted various viewpoints regarding what "problem-solving abilities" are and what they are not. At one end, one could simply want students to confidently find an answer to a problem that is both superficially and conceptually isomorphic to problems they have seen solved beforehand. At the other end, one may insist that problem-solving ability is nothing short of a student's capacity to fluidly and creatively tackle challenging and novel (to them) problems. For our part, we have adopted a pragmatic stance for this project. Our goal for this instrument is that it will prove to be a robust assessment of a student's competence to solve problems (in a restricted domain set) similar to end-of-the-chapter problems from a typical textbook. To accomplish our goal, we have settled on a narrow mechanism to characterize students' problem-solving abilities. Simply stated, on our survey students must be able to read and understand a problem statement similar to the sort one would encounter as an end-of-the-chapter problem in a typical textbook; classify the single content area relevant to that problem and identify equations germane to that area; and mathematically manipulate symbols and numbers to arrive at a correct numerical solution.

MOTIVATION

Our motivation for developing this assessment tool is two-fold. Primarily, we recognize an unmet need in the physics education instructional community. For some, such an instrument would help compliment reform efforts in other areas, since many instructors feel obliged to verify that curricular interventions in one aspect of their course, say, in conceptual understanding, leave intact their students' ability to solve problems. Other instructors are looking for a standardized way to compare their students' problem-solving abilities to other students, and they require an accepted and widely-administered instrument with an accessible and extensive data pool against which they can compare results from their courses. Whatever the case for instructors, the need to gather data characterizing students' problem-solving abilities may stem from several sources. Individual instructors may want data simply for their own reasons, or they may require it to help satisfy some external demand at departmental or administrative levels.

The second factor motivating us to create this assessment survey was the recognition of a gap in the tools at the disposal of physics education researchers. As mentioned earlier, this community has for many years investigated questions surrounding students' problem-solving abilities. Although a broadly accepted definition of "problem-solving ability" may be elusive, having a survey that at some level carefully and consistently measures, at least in part, what many consider to be skills essential to solving physics problems would be an asset for future research efforts. Once a tool is in place, researchers can begin to ask new questions or revisit older ideas and generate data to support or dismiss various perceptions about problem-solving. Also, many members of our community develop and rigorously test innovative curricular materials related to problem-solving. Possessing a tool to quantify the effects of an assortment of classroom strategies and approaches would serve those efforts, as well.

THE SURVEY INSTRUMENT

The domain of our survey instrument is intentionally limited; the three concept areas are Newton's Laws, energy, and momentum. Quite simply, these three areas represent a widely-accepted subset of important topics in a first-term physics course. We intend each item on the survey to map to one of these concepts areas. For example, a momentum problem on the survey cannot (easily) be solved using Newton's Laws. This approach will offer instructors and researchers a clearer picture of their students' problem-solving abilities in these concept areas. Furthermore, our experience has shown that single-concept problems are more than challenging enough for the majority of students, even post-instruction.

Once we settled on the domain, it naturally followed to decide on how to grade students' work. Grading can proceed along one of three general approaches: multiple-choice, open-ended grading with partial credit, or dichotomously scoring a student's final answer. All of these approaches have their advantages. Initially, we decided on a dichotomous scoring scheme because it seemed simplest, would allow for machine grading, and left open the possibility that an instructor could still grade for partial credit, if desired. Following our analysis of data from the first administration of the survey in Fall 2009, we concluded that dichotomous scoring had real analytical potential. (Please see the companion article in these Proceedings authored by Cummings and Marx for more details on data and conclusions from the first administration of this survey. [8])

Next, we considered the appropriate number of items to devote to each concept area and how the difficulty of these items should be structured. Because we want this survey to appeal to a wide range of instructors serving students with very different levels of preparation, and since we originally envisioned this survey to be administered pre-instruction and post-instruction, we felt it necessary to have a range of item difficulties for each concept area. So, we developed a Newton's Laws item, an energy item, and a momentum item that was a straightforward application of those ideas and allied equations. Then we developed another set of three items that was a bit more challenging, and so on. The cognitive load of each consecutive set of items was "ramped-up" to maximize the diagnostic potential of the data generated. We constructed successively more difficult problems through some combination of adding extraneous information, expanding to multiple dimensions, invoking misconceptions, and setting up typical mathematical missteps. The mathematics required to solve any of the items always remains at the level of basic algebra and trigonometry, and we avoided situations in which students need to employ a mathematical "trick" to solve the problem. In fact, we created the problems so a student lacking a calculator would be able to arrive at an answer without considerable computation. (We discuss the role of calculators later in this paper.) In the end we wrote fifteen items, five for each concept area.

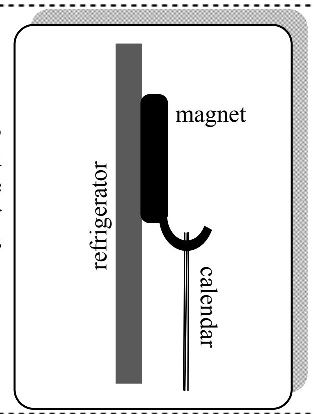
We composed each of the items such that it would look and read like a problem a student would encounter as an end-of-the-chapter problem found in most textbooks. However similar though, we invented these problems, so they should not exactly match any problem a student may have encountered. (Of course the possibility of merely coincidental previous exposure always exists.) For all but the first few items we provided an accompanying illustration related to the problem. Certainly being capable of converting a problem statement into a picture is an important skill; nevertheless, we need to limit our focus, so this is a facet of problem-solving we will not assess with our survey.

Figure 1 below displays examples of problems from version 1 of our survey. Among other things, these examples demonstrate how the first few problems of the survey are relatively straightforward, but problems toward the end of the survey will challenge most students.

3. A Volkswagen Bug (mass = 1000 kg) is traveling down the street at 5 m/s during an ice storm. The Bug approaches an SUV (mass = 3000 kg) that is stopped and unable to move due to the ice. The driver of the Bug applies the brakes, but the car just slides forward without slowing down. The two vehicles collide and stick together. How fast is the SUV moving immediately after the collision?

5. Mary Jo and Bill are rearranging furniture in their house, but they don't always agree on where all the items should go. At one instant, Mary Jo pushes a large chest of drawers (mass of 20 kg) horizontally to the right with a force of 35 N, while Bill pushes horizontally to the left with a force of 45 N on. The chest is on a slippery marble floor, so it slides with essentially no friction. (Mary Jo and Bill are both wearing rubber-soled shoes, so they get good traction.) What is the magnitude of the acceleration of the chest?

13. Dylan hangs a calendar on a refrigerator with a magnet. Unfortunately, the magnet is too weak, so the magnet and calendar slide down the side of the refrigerator to the floor with an acceleration of magnitude 3 m/s^2. If the coefficient of sliding friction between the magnet and the refrigerator is 0.1, what is the magnitude of the total force the refrigerator exerts on the magnet? The mass of the magnet is 0.05 kg and the mass of the calendar is 0.1 kg. (Ignore any contact between the calendar and the refrigerator.)



14. Two pool balls, each of mass 0.2 kg, collide as shown in the figure [to the right]. Before the collision, the black ball's velocity makes an angle of 30^o with the horizontal line. After the collision, the white ball's velocity makes an angle of 30^o with the vertical line. What is the black ball's speed after the collision?

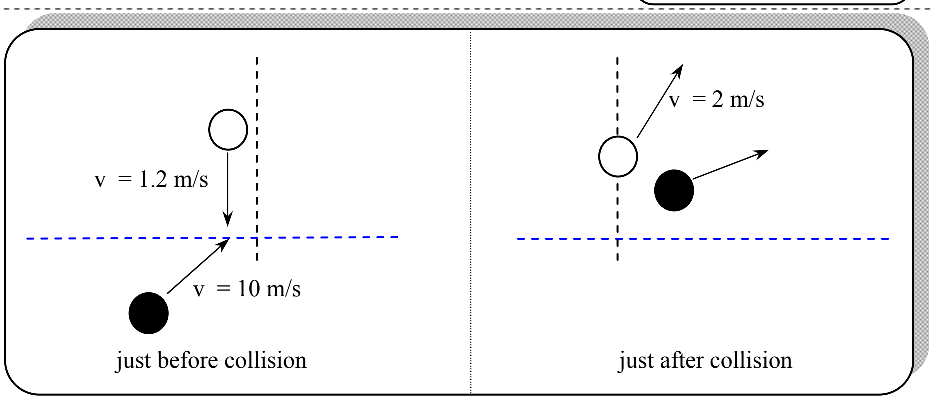


Figure 1. Four examples from the survey instrument administered in Fall 2009. The numbers here coincide with the numbering scheme on the survey. We have made some formatting modifications to fit the prescribed margins, and item 5 had a picture in the administered survey, but it was omitted here because of space considerations.

The survey instrument is designed to be presented to the students in the form of a packet. The cover sheet of the packet includes formulae and constants; a table containing values of sine, cosine, and tangent for select angles; and a right triangle with some geometric relations. At the bottom of the cover sheet is a glossary explaining the meaning of each symbol, superscript, subscript, and embellishment in the formulae. Each page of the remainder of the packet is devoted to a single item. Every item has an item number, the problem statement, and often an accompanying sketch. Many of the problem statements also include a reminder to the examinee that the values of sine, cosine, and tangent for specific angles are listed on the cover sheet. Immediately below all of that is a line and a statement directing students to put all of their work below the line and to write their final answer in the box at the right-hand side of the sheet. As we stated earlier, the first three items were, in our estimation, the simplest three, covering Newton's laws, energy, and momentum. The next three items were slightly more demanding, and so on. We arranged the three items in each successive set such that they did not follow any obvious pattern.

Students are also provided written instructions at the beginning informing them that they are not to use calculators, but if needed they could "leave any numerical answer in a convenient (yet sensibly simplified) form, such as ⅞ or $\sqrt{8.3}$." Also, they could use any of the formulae from the cover sheet in addition to any they knew already. And, they are to ignore the effects of air resistance and let the local gravitational field strength be approximated by 10 N/kg = 10 m/s^2.

At the start of the administration, the instructor informs the students that they have forty-five minutes to complete the survey, may work on the problems in any order, are not permitted to use calculators, and that final answers need to be written in the appropriate box to be considered by the grader.

For the first version of the survey we decided to not permit students to use a calculator, and so we developed problems in which the calculations "worked out." We felt the use of calculators by the students was problematic for a few reasons. First, we felt that if faculty wanted to administer the survey unannounced and it required the use of a calculator, then some students would be at a disadvantage. Second, it has been our observation that when students have access to the use of a calculator they spend a significant amount of time working with the calculator and not thinking about their solution. Finally, some instructors simply do not permit their students to use calculators, so developing an instrument required accommodating their concerns.

As we stated above, we planned for students to have forty-five minutes to work. We deemed this an appropriate length for a few reasons. First, the few physics instructors on whom we tested an early version of the survey were able to finish it in about that length of time. If that seems as though we have then been unfair in allotting time for students, it is essential to bear in mind that we do not anticipate many students will be able to finish this survey within the time constraints. It is our goal that only the most efficient and prepared students will answer every item. Regardless of the final length and time limit of the exam, we expect the survey to tax the abilities of nearly all students in our target population. Related, we know that a survey requiring more time would make widespread adoption less likely. Instructors have limited time; expecting them to devote more than forty-five minutes to a survey instrument would be burdensome. (Even forty-five minutes seemed excessive; we envision a final version to have something closer to a thirty-minute administration period.)

Finally, it seemed important to develop an instrument that could be administered pre-instruction and post-instruction. Clearly many instructors and researchers would find enhanced utility in an assessment tool intended to measure shifts in students' abilities to solve problems. So, when possible we tried to use everyday language while wording our items. And, as we mentioned before, the students were provided with a crib sheet that contained all of the equations they would need to solve any of the problems along with some explanation of each of the terms and symbols.

LOOKING AHEAD

At the heart of this project is our desire to develop an instrument that physics instructors and education researchers will find useful. This quest for broad acceptance has helped us look for middle ground in a wide field of interests and concerns, and many of the issues outlined in this paper are still very much under consideration. Clearly the administration length and number of items is of serious concern. Matters regarding appropriate grading schema and the use of calculators require further consideration, too. Still other issues were not debated at the outset, but came to our attention as the survey was administered in the Fall 2009. In particular, we speculated about whether or not we should recommend that instructors could count this survey as part of a student's course grade, or if the relevant population of examinees could be extended to include students in an algebra-based course. We have begun investigating some of these issues through analysis of data from our first administration in Fall 2009. Again, we encourage people to please read the companion article mentioned earlier.

We close by inviting instructors interested in administering our survey to contact either one of the authors at the e-mail addresses listed below.[9]

REFERENCES AND NOTES

1. F. Reif, *Phys. Teach.* **19**, 310-316 (1981).
2. A. Van Heuvelen, *Am. J. Phys.* **59**, 891-897 (1991).
3. J. I. Heller and F. Reif, *Cogn. Instruct.* **1**, 177-216 (1984).
4. M. T. H. Chi, et al. *Cogn. Sci.* **5**, 121-152 (1981).
5. D. P. Maloney, *J. Coll Sci. Teach.* **16**, 510-514 (1987).
6. J. H. P. van Weeren, et al. *Am. J. Phys.* **50**, 725-732 (1982).
7. P. Heller, et al. *Am. J. Phys.* **60**, 627-636 (1992).
8. K. Cummings and J. Marx, *PERC 2010 Proceedings (2010)*, in press.

9. *Contact e-mail information*

Jeff Marx: jmarx@mcdaniel.edu
Karen Cummings: cummingsk2@southernct.edu