Assessment of Student Problem Solving Processes

Jennifer Docktor and Kenneth Heller

School of Physics and Astronomy, University of Minnesota, Minneapolis, MN 55455

Abstract. Problem solving is a complex process important both in itself and as a tool for learning physics. Currently there is no standard way to measure problem solving that is independent of physics topic, pedagogy, and problem characteristics. At Minnesota we have been developing a rubric to evaluate students' written solutions to physics problems that is easy to use and reasonably valid and reliable. The rubric identifies five general problem-solving processes and defines the criteria to attain a score in each: useful description, physics approach, specific application of physics, math procedures, and logical progression. An important test of the instrument is to check whether these categories as represented in students' written solutions correspond to processes students engage in during problem solving. Eight problem-solving interviews were conducted with students enrolled in an introductory university physics course to compare what students write down during problem solving with what they say they were thinking about as determined by their interview statements.

Keywords: problem solving, rubric, assessment
PACS: 01.40.Fk, 01.40G-, 01.40.gf, 01.50.Kw

INTRODUCTION

Although problem solving is widely believed to be an important part of most physics courses, there is no standard measurement of student proficiency in such processes. This makes it difficult to determine the effectiveness of curricular materials or pedagogies. The goal of this study is to design and test a rubric to assess written solutions to physics problems.

The framework for this study is based on educational measurement theories of validity, reliability, and utility and is explained elsewhere [1]. Validity refers to the extent to which an assessment's interpretations are supported by theory and empirical evidence. Sources of evidence for validity come from multiple aspects of an assessment including the relevance of its content, its consistency with response processes of examinees, and its generalizability [2]. The problem-solving interviews described in this portion of the study focus on the evidence of its fidelity to student response processes.

The rubric developed at Minnesota identifies five general problem-solving processes and defines the criteria to attain a score in each. Although problem solving is very complex, these processes give a broad representation of major aspects and were chosen to be as congruent as possible to the expectations of physics instructors, research literature on problem solving, and the structure of written solutions. *Useful Description* refers to the process of summarizing information from a problem statement in an appropriate and useful form, such as assigning mathematically useful symbols to quantities and visualizing the situation with a sketch. *Physics Approach* is the process of selecting appropriate physics concepts and principles for the problem and having a basic understanding of those concepts. *Specific Application of Physics* is the process of linking concepts and principles to the specific quantities and assumptions in the problem. *Mathematical Procedures* refer to the mathematical operations used to obtain the desired physics quantity. *Logical Progression* is an overall category that assesses the extent to which the solution is focused and consistent. Scores on the rubric range from 0-5 with additional Not Applicable (NA) categories of processes unnecessary for a problem (NA Problem) and aspects missing but could be unnecessary for the solver to write down explicitly (NA Solver).

Problem solving interviews were conducted with introductory physics students to investigate the extent to which these five rubric categories represent the written and unwritten processes students engage in during problem solving. The analysis addresses:

- How is what students write down during a problem solving interview similar to or different from what they say they thought about?
- To what extent are the rubric category processes observed during a problem-solving interview?
- What problem-solving processes are observed that are not explicitly measured by the rubric?

DATA COLLECTION & ANALYSIS

Participants in the problem-solving interviews were students enrolled in an introductory calculus-based mechanics course for scientists and engineers in the Spring 2009 semester. Of the 238 students in this course, 13 volunteered to participate in a one-hour problem-solving interview at the end of the term. Ten of these students were eligible to participate and scheduled a session time. Four interviews took place in the last week of the semester, 2 during finals week, and 2 after finals week. Two students cancelled giving a total of 8 interviews: 7 males and 1 female.

The final course grades of these eight participants indicate they performed higher than the course average and may not accurately represent the problem solving proficiency of their class. Students 2, 5, 6, and 8 received an A in the course; Students 3 and 4 received an A-; Student 1 earned a B+; and Student 7 a B. There was also a higher fraction of non-native English speakers participating in the interviews (two of eight)

During the problem-solving interviews students were asked to work on physics problem(s) while being video and audio taped. They used large sheets of paper and a black marker to record their solution(s). Participants were asked to talk out loud while working on a problem if that was comfortable for them, or they could wait and explain their solution at the end. Only one student (Student 1) opted to talk out loud while working. Students were provided a copy of the instructor's equation sheet from the course and their calculator.

The problem tasks were selected to look similar to ones from tests and group problem-solving sessions in their course (*context-rich* problems) [3]. They were written to require several decisions from the solver and did not include any illustrations. The first and most involved task is given in Figure 1. This problem was adapted from previous research [4]. Problem features include: the target of the problem is not explicitly stated, a combination of at least two principles is necessary, and the solver must infer or assume some information. This problem also has the characteristic that it is possible to obtain a correct answer with incorrect or incomplete reasoning.

The remaining two problems were designed to be shorter (in anticipation of little available time). These problems maintained the same context-rich format but only required a single physics principle. Problem two required finding the spring constant of a bungee cord and Problem three involved a car crash at the bottom of a cliff. Students were only given an additional problem if sufficient time (at least twenty minutes) remained after they had explained their reasoning for the previous problem.

You are working at a construction site and need to get a 14-N bag of nails to your co-worker standing on the top of the building (9 meters from the ground). You don't want to climb all the way back up and then back down again, so you try to throw the bag of nails up. Unfortunately, you're not strong enough to throw the bag of nails all the way up so you try another method. You tie the bag of nails to the end of a 65-cm string and whirl the string around in a vertical circle. You try this, and after a little while of moving your hand back and forth to get the bag going in a circle you notice that you no longer have to move your hand to keep the bag moving in a circle. You think that if you release the bag of nails when the string is horizontal to the ground that the bag will go up to your co-worker. As you whirl the bag of nails around, however, you begin to worry that the string might break, so you stop and attempt to decide before continuing. According to the string manufacturer, the string is designed to hold up to 500 N. You know from experience that the string is most likely to break when the bag of nails is at its lowest point.

FIGURE 1. The first problem-solving task presented to interview students.

After solving a problem to their satisfaction, each student was asked to go back and explain their solution to the researcher. Questions from the semi-structured interview included the following:

- When you read through the problem, what was the first thing you thought about?
- What did you think about next?
- What was the first thing you wrote down?
- What did you think this question was asking you to find?
- How did you decide to use ___ ? (physics)
- If you were solving this problem on an exam, what would you hand in to be graded?
- Have you solved a problem like this before in your physics class? How is that problem similar to or different from this problem?
- While you were working on the problem, was there anything you did in your head that you didn't write down?

The audio files for the eight interviews were transcribed and the written protocols were analyzed using Q.S.R. NVivo® software using fourteen prescribed code categories or "nodes". Five of these nodes corresponded to the process categories on the rubric and eight others designated specific questions asked during the interview (stated above). There was also an "Other" code to identify processes observed that were not explicitly addressed by the rubric.

RESULTS

The average time each student spent working on the first problem is listed in Table 1, along with the total number of problems completed during the interview session. The times ranged from 6 minutes to 26 minutes, excluding additional time spent modifying the solution during questioning. For the three students who solved it most quickly (Students 2 and 5) it is possible that the task was not a problem for them but an exercise. Students 4 and 7 did not reach a satisfactory answer for the problem and chose to stop at the reported time to explain their thinking.

TABLE 1. Time spent working on first problem and total number of problems completed for each interview student

Interview Student	Time Spent on First Problem	No. Problems Completed
Student 1	14 min 55 sec	2
Student 2	6 min 50 sec	3
Student 3	24 min 30 sec	1
Student 4	26 min 17 sec (+13 min 5 sec)	0
Student 5	6 min 7 sec (+ 40 sec)	2
Student 6	14 min 54 sec	2
Student 7	20 min 27 sec (+ 5 min 33 sec)	0
Student 8	9 min 56 sec	2

(Times in parentheses indicate that students changed or added to their solution during the interview questions)

Student 2 was the only student to successfully complete the problem with correct physics reasoning. Student 5 was successful after correcting an error discovered during the interview questioning. Students 6 and 8 obtained the correct answer, but gave incomplete (possibly incorrect) reasoning for some quantities in their equations. Students 1 and 3 completed the problem using inappropriate physics equations. Students 4 and 7 did not obtain a final answer in the available time and their approaches included a mixture of confused physics ideas. Two students interpreted the question as finding the height the bag would travel vertically with the maximum string tension value, whereas most students focused on solving for a force. Student four's goal was unclear.

All eight students began by writing down a diagram of the problem situation and summarizing the information provided. When asked what they thought about while reading the problem, three students (1, 6, and 8) mentioned this problem description process:

S8: *Well the first thing I uh, thought about was um…I just diagrammed it. I didn't know what to think initially. I just wrote down all the data, diagrammed it.*
Q: *Okay, when you say 'diagrammed' can you tell me more of what you mean by that?*
S8: *Like, I just like, visualized it. Maybe the height had to be from the center of the, center of the thing. I wasn't quite sure exactly what it was but when I drew a picture it made more sense to me.*

Student 5 said they first thought about what the question was asking them to find. Students 3 and 4 mentioned that the problem made them think about circular motion, and Student 7 mentioned parabolic motion because it was like "throwing something". Student 2 explicitly mentioned physics principles:

S2: *That it is not hard…and I should use uh, the equation of the motion and uh, the conservation of energy in this problem.*

The total number of transcript passages assigned to each coding category is reported in Table 2, in addition to the average number of passages for students 2 through 8. As seen from Total column, most statements pertaining to the rubric categories (the first five nodes) were coded as evidence of specific application of physics or logical progression. Specific application statements were usually references to particular physics equations and the quantities specific to the problem (such as a velocity, force, or distance). For example, when prompted students 6 and 8 stated the velocity of the bag was the same at all points of the swing, which was not obvious from their written work. Logical progression statements referred to overall steps taken in the solution and explaining reasoning for those steps. The following statement from Student 5 is an example of this category.

S5: *Um, first uh, I find out what I want to know. And I find out what I already know. And I need to build a relationship between them… in this problem I want to know the height so I need to know the velocity. And in order to find the velocity I need to know the, use Newton's second law I can find the, the relationship between the force and the velocity. So I build the connection with the known things and the other things.*

In contrast, Student 4 describes their procedure:

S4: *Pretty sure I'm lost.*
Q: *Can you say more about that? What are you, what are you thinking right now?*
S4: *I can't really, I don't really know. I was just trying to put everything I know down, and then seeing what equations eliminate stuff. Um, and what I could plug in. And that didn't get me very far so far.*

There were 549 total passages coded in the eight interview transcripts. On average, students made 32 rubric-related statements of 65 total statements coded. Student one talked out loud while solving the problem and had more: 55 rubric-related statements and 95 total statements. Of the statements in the "Other" rubric category, half pertained to monitoring progress, evaluating the answer, and/or checking units. Although these processes are desirable and could contribute to the logical progression of a solution they are not explicitly scored by the rubric. Additional processes in this category included solving equations in symbolic form prior to plugging in numbers and referencing the equation sheet.

Students who spent a lot of time on the problem (3, 4, and 7) generally had a higher number of statements about deciding what physics to use and engaging in mathematical procedures, because they attempted several different approaches.

TABLE 2. Number of transcript passages assigned to each coding node

Coding Category (NVivo Node)	Total Passages	Avg. Passages (S2-S8)*
Useful Description	43	5
Physics Approach	40	4
Specific Application	82	9
Math Procedures	39	5
Logical Progression	72	9
Other	84	10
Thoughts First	17	2
Thoughts Second	12	2
Write Down First	17	2
Decide What to Find	30	4
Decide What Physics	30	3
Write Down on Exam	17	2
Previous Problem	22	3
Processes in Head	44	5
TOTAL	**549**	**65**

*Student 1 talked aloud while working on the problems and generally had more statements coded than the other students.

When asked what they would hand in for a graded exam, all students gave examples of adding more explanation in words to help the grader understand the solution, and most said they would write a picture.

S6: *Um, I would start with two diagrams at the top kind-of. Showing all of this basic information. And then…I would kind-of explain maybe in a phrase or something what each of these different sections were doing, and I'd kind-of put them in a logical order as opposed to here where they're, it's a little bit um, jumping all over the page…just so that it's clear.*

When explicitly questioned about what they did in their head and didn't write down, Student 7 described:

S7: *Usually the only thing I write down right away is a picture, so I can see what's going on. Um. But then I'll just have in my head like, if I go from this equation and then I get an answer I can put it into this equation, and then into that equation…Generally I tend to do too much in my head and not write enough stuff down, that's the only, that seems to be where I go wrong.*

SUMMARY

In this study eight introductory university physics students each participated in a one-hour interview to compare their written and unwritten problem solving processes. For the first problem presented to them, all students wrote down a description, physics equations, and mathematical operations. However, during the interview they said they thought about processes such as interpreting the question, planning steps, deciding among multiple possible physics concepts, considering previously solved problems, determining whether they should abandon their approach and try something else, and evaluating the reasonableness of their answer.

The transcripts contain evidence for all five of the rubric categories, with specific application of physics and logical progression having the most coded statements. Since students were prompted to explain their reasoning verbally during the interviews, evidence of logic was much more prominent than is typical of their written work. Also, there was explicit evidence for the physics approach (selecting appropriate physics principles) that is often inferred from the specific equations that are written down. In general, rubric scores of students' written solutions alone were consistent with the verbal evidence. An exception was Students 6 and 8, whose incorrect physics reasoning was not apparent from their papers. This affected only their *Specific Application of Physics* category from X to Y.

In summary, the process categories of the problem-solving rubric are observed in both written work and verbal interview protocols. There is, of course, much more fine-grained information in the interviews. Also, according to the students, much of what they hand in on a test is a "cleaned up" version of a problem solution that may contain more information than their interview papers. From these interviews, we conclude that rating student solutions using a problem solving rubric generally gives an accurate, though course-grained, view of their problem solving processes.

ACKNOWLEDGMENTS

This research has its origins in work given in the dissertations of Tom Foster and Jennifer Blue and is supported by NSF grant DUE-0715615.

REFERENCES

1. J. Docktor and K. Heller, "Robust Assessment Instrument for Student Problem Solving" in *Proceedings of the 82nd NARST Annual International Conference* (171 April 2009).
2. AERA, APA, and NCME, *Standards for Educational and Psychological Testing*, Wash., DC: AERA (1999).
3. P. Heller and M. Hollabaugh, *Am. J. Phys.* **60**, 637 (1992).
4. C. Henderson, E. Yerushalmi, V. Kuo, P. Heller, and K. Heller, *Am. J. Phys.* **72**, 164 (2004).