Problem Solving Skill Evaluation Instrument – Validation Studies

Wendy K. Adams and Carl E. Wieman

Department of Physics, University of Colorado, Boulder, Boulder, CO 80309, USA

Abstract. Researchers have created several tools for evaluating conceptual understanding as well as students' attitudes and beliefs about physics; however, the field of problem solving is sorely lacking a broad use evaluation tool. This missing tool is an indication of the complexity of the field. The most obvious and largest hurdle to evaluating physics problem solving skills is untangling the skills from the physics content knowledge necessary to solve problems. We are tackling this problem by looking for the physics problem solving skills that are useful in other disciplines as well as physics. We report on the results of a series of interviews comparing physics students' skills when solving physics problems with their anonymous completion of the problem solving instrument. There is an encouragingly good match.

Keywords: Problem Solving, Physics, Undergraduate Education
PACS: 01.40.Fk, 01.50.gf

INTRODUCTION

Physics Education Researchers over the past 20 years have developed many very useful assessment tools including a range of conceptual evaluations such as the Force Concept Inventory [1], Force and Motion Conceptual Evaluation [2], and the Conceptual Survey of Electricity and Magnetism [3] among others. There has also been work in the area of attitude and belief surveys [4]. However, one of the most highly valued assets of physics courses by educators is problem solving; but, very little has been published regarding ways to evaluate students problem solving skills. To date, research on problem solving in physics has been focused predominantly on curriculum improvements and the identification of differences between expert and novice problem solving characteristics [5]. There have been efforts to develop a problem solving evaluation tool, but none have been successful.

In this paper we will discuss the problem solving instrument that is being created at Colorado. First we will describe the direction of our approach and how it avoids problems that have hindered previous attempts. Then we will outline our validation procedures to date, and present the specific set of measurable skills that we have identified.

APPROACH

First we define what we mean by "problem solving", because this term is used very broadly. The fairly specific definition of problem solving used for the purposes of this paper is drawn from the literature: "Problem solving is cognitive processing directed at achieving a goal when no solution method is obvious to the problem solver." [6]. In addition we will break problem solving down into a set of skills, including those motivational aspects relevant to successful completion of solution. Note that this definition of problem solving is based on the solver. If a person is an 'expert' in their field, then it is very likely that tasks that are problems for students will only be an 'exercise' where the solution or path to solution is obvious, for an expert.

The extensive problem solving research in psychology and cognitive science has included transfer of problem solving ability as well as work on characterizing specific skills needed to solve problems. Many researchers have gone as far as to declare that problem solving abilities do not transfer between disciplines [5]. Being an expert problem solver in a particular area is not an indication of that person's problem solving skills in any other area. However, other researchers present evidence that some problem solving skills do transfer between disciplines [6].

We argue that these contradictory results on problem solving transfer can be explained with the same reasons that have made developing a physics problem solving evaluation tool difficult. First, solving physics problems requires content knowledge, both factual and procedural; second, that it is important to view problem solving as a combination of many different specific skills, and the different skills have varying degrees of importance in various types of problems.

The physics content requirement is a fairly straight forward yet seemingly unassailable complication. A particular student may be unsuccessful at solving a problem because they are a poor problem solver or because they simply lack familiarity with the specific bit of physics knowledge needed for the problem. On the other hand, when a student is successful at solving a particular problem, it is very difficult to tell the difference between a strong problem solver or a student who is so familiar with that particular content area that the problem was actually an exercise for that student. In both of these cases, one cannot clearly delineate the student's problem solving ability because the student's content knowledge is inextricably intertwined within these skills.

A great deal of research in problem solving – physics, chess, etc. is result oriented. Researchers are interested only in who can successfully solve problems rather than each solvers' strengths and weaknesses in a variety of specific problem solving skills. This focus greatly limits both the information that can be gained by these studies and the development of assessment tools.

Our approach comes from the hypothesis that a person has a set of skills that vary in strength that they use to tackle problems. This includes problems in any context, the physics classroom or the workplace. If this is the case, students' skills can be analyzed with a problem that does not require physics content knowledge. Although still preliminary, we have growing evidence that our hypothesis is correct.

ASSESSMENT TOOL

We have developed an assessment tool which uses the work of the Cognition and Technology Group at Vanderbilt (CTGV)[7]. CTGV developed the Jasper Woodbury series of problems for 6^{th} – 7^{th} grade math students to solve in small groups. These are long, involved problems that each have no less than 14 steps to solution. CTGV carefully designed and researched each Jasper problem on its effectiveness. Our problem solving instrument was developed through a series of interviews with a wide range of subjects using one of the Jasper problems. The evaluation instrument instructs the student to analyze a script of two individuals working through the solution to the Jasper problem. This format provides several benefits: 1) motivation for students to work through the entire solution, 2) removes the stress of being analyzed 3) scaffolds the problem so that a solution will be reached even if the student has a specific weakness that would have prevented further progress if they were attempting to solve the problem in isolation. There are 57 questions that must be answered as they go through the script. These questions are about planning, procedures, calculations, reflection as well as analysis of the two scripted solvers' skills. This combination of questions evolved from multiple interviews that were used to evaluate students' skills as they solved the scripted problem. In this way, we have identified 34 skills which show up as useful when solving this problem.

VALIDATION

We have performed a series of validation studies that involved several iterations to revise and refine the instrument. The validation process included: face validity – interviews with a wide range of people to confirm and clarify the meaning of questions and the story; anonymous written instrument results compared to a series of physics problem solving interviews; and concurrent validity – comparison of instrument results to professor and employer evaluations. In addition we have two studies which are in process: interviewer observations of physics majors using the instrument will be compared to an independent interviewer's observations of the same students solving quantum mechanics problems; and written instrument responses of undergraduate teaching assistants will be compared to independent instructor evaluation of the teaching assistants using the same rubric.

The development of the survey was intertwined with the first phase of face validation. Interviews were conducted with a variety of subjects including non-science majors, physics majors, professors, elementary education majors, and adult professionals with a variety of backgrounds including a high school drop out. The first few interviewees were asked to think out loud as they solved the Jasper problem. The two person script and the questions that are asked throughout the instrument were created based on these first few interviews. Further interviews were conducted with the scripted instrument. Periodically the script was refined and questions were added based on interviews. A total of 23 interviews were completed, each lasting between one and two hours.

Eventually the instrument and analysis rubric were refined to the point that for the last nine interviewers, the interviewer was able to obtain a complete analysis of the subjects skills without further interaction after initial instructions on think-aloud style. The analysis rubric includes 34 separate problem solving skills that have been identified during the interviews (see Table 1).

The next stage of instrument construction and validation of the rubric involved having the students take a written version (no interview) of the survey. These were then graded using the same rubric. If enough information was not provided in the written responses to rate the student in all 34 categories, then questions were added or adjusted and tested with new students. Currently 16 written responses have been graded and 90% of the skills that are identifiable during interviews can be consistently graded with the latest written version of the survey.

Anonymous Written Results Matched to Physics Problem solving Interviews

The next step in face validation was to test how the set of problem solving skills that are identified while solving this 6^{th} grade math trip-planning problem matched with the skills a student uses to solve a physics problem. For this portion of the validation procedure we gave five science majors who were currently enrolled in introductory physics the written version of the instrument. These students brought the completed instrument to their first interview. The instrument was graded at a later date without knowing the identity of the student. Interviews consisted of having the student sit down with the interviewer whom they had never met and immediately begin solving a slightly modified problem about the Great Pyramid of Giza[7]. In this problem students must determine how many blocks are in the pyramid and how many men were needed to build the pyramid if it took 20 years to build. A few facts about the block size and men's capabilities were included without any other scaffolding. These interviews were mathematically intensive so had to be limited to an hour. Each student required two to three interview sessions to complete the problem. They were evaluated on the skills that

TABLE 1. Problem Solving Skills

Skills	**Metaskill**	**Will**	**Uncatergorized**
Skills that were measured by both the written instrument and the pyramid interview			
◆Math Skills (add/sub/mult/div)	◆Planning What – Question Formation	◆Confidence	◆Real Life vs. Classroom Approach
◆Spatial – Mapping	◆Planning How – Way to get answer.	◆Enjoyed Solving the Problem	◆Overall Success
◆Estimation	◆Planning Big Picture – Visualizes the Problem.	◆Wanted to Succeed on 'Test'	
◆Number Sense	◆Connects Steps and Pieces	◆Attribution (responsible for own mistakes)	
◆Real World (informal) Experience	◆Monitors Own Progress	◆Wanted to Find Best Solution for Themselves.	
	◆Knowledge of Own Strengths and Weaknesses	◆Wanted to find Best Solution to Please Interviewer.	
`	◆Creativity		
	◆Judgment of Reasonable Issues		
	◆Ties in Personal Experience		
Skills that were not measured by the pyramid problem interviews and why			
◆Acquires Information 1st Time Through (*pyramid data of a numerical nature)*	◆Keeps Problem Framework in Mind (*Pyramid framework not complicated)*	◆Enjoyed Analyzing Interns (*Specific to instrument scenario)*	
◆Remembers Previously Noted Facts. (*pyramid interviews done in 2 segments w/ a 2 week break.)*	◆Adaptability - Shift Direction (*A forced change of plans was not inserted into Pyramid problem)*	◆Enjoyed Complete Survey (*Same as enjoyed solving the problem for pyramid)*	
◆Outside Factual Knowledge (*Not necessary for pyramid problem or in separate category such as Geometry)*	◆Skepticism – Thinks about Information that is Supplied. (*Specific to instrument scenario)*		
◆Math – Equation Formation (*Difficult to see w/ Pyramid response format)*	◆Checks Scripted Solvers' Calculations (*Specific to instrument scenario)*		
◆Reading Comprehension (*Pyramid presentation makes this difficult to judge*)	◆Aware of How Scripted Solvers Helped (*Specific to instrument scenario)*		
Skills that were not measured by the written problem solving instrument and why			
◆Geometry (*Not necessary for Survey)*	◆Physics/Math They Think They 'Should' Know Blocks Progress (*No outside formal knowledge needed for survey)*		

were exhibited while completing the problem. The interview sessions were separated by two weeks which made some of the grading of skills difficult, as noted in Table 1.

The five graded anonymous written instruments were then compared to each student's set of skill ratings identified during the pyramid interviews. The skills that could be identified with both the anonymous written instrument and the pyramid interviews are listed in Table 1. There were a few skills that only worked in one situation or the other. These are listed in Table 1 with explanations. Many of the skills identified by the instrument that could not be graded by the pyramid interviews have been recognized as important skills when solving physics problems, they simply could not be rated with the structure of the pyramid interviews.

The five students' written instruments were easily matched up with their interview results. This was done without any identifying information about the students other than the rating of problem solving skills from each problem scenario. The same set of problem solving strengths and weaknesses were evident in how each student solved these two different problems, After the students written results were matched with their interview results, the specific ratings in each of the skills matched up for 75% of the 22 skills listed in Table 1. Of the 25% that did not match exactly, most were a neutral response versus (strong/weak). Only two of the students had a single skill that appeared strong for one problem and weak on the other.

Concurrent Validation

Our concurrent validation studies include comparison between the instrument results and evaluation of problem solving skills in various environments. We asked three different instructors (class sizes of 40 or less) to evaluate particular students' strengths and weaknesses over the course of a semester. These same students were interviewed or took the written version of the problem solving evaluation tool. The instructor's assessment of the students strengths and weaknesses in problem solving have matched the skills identified by instrument results in all 15 cases. In addition we had an employer to do the same for five different employees. Again, all five instrument results matched the employer's evaluations.

We are in the process of conducting an additional validation study involving the comparison of instrument results with skills identified while students solve quantum mechanics problems. Six students underwent a series of seven interviews where they solved quantum mechanics problems. The interviewer is in the process of scoring these students quantum mechanics problem solving skills on the 34 areas identified in Table 1. A separate interviewer is interviewing these same students as they solve the problem solving evaluation instrument. The results of these separate interviews will be compared.

CONCLUSION

We have created an evaluation instrument that identifies 34 specific problem solving skills. We have evidence to support our hypothesis that a particular student has the same strengths and weaknesses when solving a complicated trip planning problem as they do when solving a physics problem or performing in the work place. There are a few specific skills that may be necessary for one type of problem that are not required to solve another; however, the strengths and weaknesses of a particular person are the same, regardless of the environment. Consistent strengths and weaknesses match intuition – certain people are stronger problem solvers – however; on the surface this may seem to disagree with research on context and transfer. A closer look at problem solving shows that our results are in fact consistent with other research on problem solving as well. The difference is that most research is complicated with the need for specific content knowledge to solve a particular problem and many researchers focus on the end result rather than looking at the specific skills used by the solver during the problem solving process.

ACKNOWLEDGMENTS

Supported in part by the NFS Distinguished Teaching Scholar.

REFERENCES

1. D. Hestenes, M. Wells and G. Swackhamer, *The Physics Teacher* **30**, 141-158 (1992).
2. R. K. Thornton and D. R. Sokoloff, *American Journal of Physics* **66**, 338-352 (1998).
3. D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke and A. Van Heuvelen, *Am. Jrnl. of Phys.* **69**, S12-S23 (2001).
4. W. K. Adams, K. K. Perkins, M. Dubson, N. D. Finkelstein and C. E. Wieman, *Phys. Rev. Special Topics – PER* **2** (2006).
5. D. P. Maloney, "Research on Problem solving: Physics," in *Handbook of Research on Science Teaching and Learning* edited by D. L. Gabel, Macmillan, Toronto 1993, pp 327-354.
6. R.E. Meyer, *Thinking, problem solving, cognition (2nd ed)* Freeman, New York, 1992.
7. The Cognition & Technology Group at Vanderbilt, *Ed. Tech. Rsrch. and Dev.* **40**, 65-80 (1992).
8. K. Vick, E. Redish, and P. Cooney, Retrieved 8/18/06 from Activity-Based Physics (ABP) Alternative Homework Assignments (AHAs) Problem site: http://www.physics.umd.edu/perg/abp/aha/pyramid.htm
9. A copy of the problem solving instrument can be viewed at: http://cosmos.colorado.edu/CPPSS/