Robust Assessment Instrument for Student Problem Solving

INTRODUCTION

Problem solving skills (qualitative and quantitative) are a primary tool used in most physics instruction. Despite this importance, a reliable, valid, and easy to use quantitative measure of physics problem solving does not exist. Although scoring tools have been used in past problem solving research at the University of Minnesota and other places, these instruments are difficult to use and require a great deal of time and training.

The goal of this study is to develop a robust, easy to use instrument to assess students’ written solutions to physics problems, and determine its reliability and validity. In addition, this research will necessarily develop materials for its appropriate use and training.

Validity in this context refers to the degree to which score interpretations are supported by empirical evidence and theoretical descriptions of the process of problem solving. Reliability refers to the stability of scores across multiple raters. Modern views of validity theory focus on collecting different kinds of evidence and considering the appropriateness, meaningfulness, and usefulness of a performance measure.

CATEGORY DESCRIPTIONS

The instrument takes the form of a grid or rubric, which divides physics problem solving into five sub-skill categories. These sub-skills are based on those identified by research in cognitive psychology, especially the investigations of the differences between expert and novice problem solving processes.

USEFUL DESCRIPTION

Organize information from the problem situation symbolically and visually

PHYSICS APPROACH

Select appropriate physics concepts and principles

SPECIFIC APPLICATION OF PHYSICS

Apply physics approach to the specific conditions in the problem

MATH PROCEDURES

Follow appropriate and correct mathematical rules / procedures

LOGICAL PROGRESSION

The solution progresses logically; it is coherent, focused toward a goal, and consistent

PILOT STUDY

• In Fall 2007, eight experienced graduate student teaching assistants used the rubric to score students’ written solutions to final exam problems before and after a minimal training exercise. They also provided written feedback on the rubric categories and scoring process.
• Four volunteers scored mechanics problem solutions and four scored E&M solutions.
• Before training 8 solutions were scored. Training consisted of example scores and rationale for the first 3 solutions. Five solutions were re-scored, and 5 new solutions were scored.

RATER AGREEMENT

Weighted Kappa Measure of Agreement:

\[ \kappa = \frac{P_o - P_e}{1 - P_e} \]

f\(_o\): observed frequencies
f\(_e\): expected frequencies above chance
w: weight coefficient (square of the difference in two raters’ scores)

EXAMPLE STUDENT SOLUTION

To raise money for a University scholarship fund, the dean has volunteered to bungee jump from a crane. To add some interest, the jump will be made from 42 m above a pool of water. A 30 m bungee cord would be attached to the dean. First you must conceive the dean that your plan is safe for a person of his mass, 75 kg. You then have the dean step off a platform and being in free fall 30 m before the cord begins to stretch. Determine the spring constant of the bungee cord so that it stretches only 12 m, which will just keep the dean out of the water. Find the dean’s speed 7 m above the water.

TRAINING

REVISIONS AFTER PILOT

• The scoring scale was increased by 1. The former “0” score was separated into two, one for all inappropriate and one for all missing.
• The NA(Problem) and NA(Solver) categories were included more prominently in the rubric.
• The Useful Description category was moved before Physics Approach because symbolic and visual descriptions usually appear first in a solution.
• The wording was made more parallel in every category.

REVISED PROBLEM SOLVING RUBRIC

<table>
<thead>
<tr>
<th>Category</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>NA(Problem)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>NA(Solver)</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Physics Approach</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Specific Application of Physics</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Mathematical Procedures</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Logical Progression</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>7</td>
</tr>
</tbody>
</table>

FINDINGS

• Rater agreement with the researcher was poor to fair before training and improved to fair or moderate agreement after a minimal training.
• Perfect agreement was 34% before training and increased to 44% after training. Agreement within one score was 77% before and 80% after training.
• After training, raters’ scores for some categories decreased (especially Math and Logic) to match the example scores.
• NA categories and the score zero were largely ignored or avoided, even after training.
• “I am confused by the need for NA(Solver). What is an example of when this would be an appropriate score?” [TA#4]
• “[The student] didn’t do any math that was wrong, but it seems like too many points for such simple math” [TA#8]

• Weighted Kappa Measure of Agreement 13:

\[ \kappa = \frac{P_o - P_e}{1 - P_e} \]

• Compare scores from the rubric with another measure of problem solving (validity measures).

• Example student solution

EXAMPLE STUDENT SOLUTION

To raise money for a University scholarship fund, the dean has volunteered to bungee jump from a crane. To add some interest, the jump will be made from 42 m above a pool of water. A 30 m bungee cord would be attached to the dean. First you must conceive the dean that your plan is safe for a person of his mass, 75 kg. You then have the dean step off a platform and being in free fall 30 m before the cord begins to stretch. Determine the spring constant of the bungee cord so that it stretches only 12 m, which will just keep the dean out of the water. Find the dean’s speed 7 m above the water.

NEXT STEPS

• Revise the training materials to include a description of the rubric’s purpose and a greater range of score examples, especially for NA scores.
• Re-test the revised rubric and training materials with graduate students and faculty to assess reliability.
• Compare scores from the rubric with another measure of problem solving (validity measures).

REFERENCES


http://groups.physics.umn.edu/physed