Robust Assessment Instrument for Student Problem Solving

Jennifer L. Docktor

Kenneth Heller, Patricia Heller, Tom Thaden-Koch, Jun Li, Jay Dornfeld, Michael Forte

Physics Education Research & Development Group
http://groups.physics.umn.edu/physed

More Details  Poster: PST4-25  5-6 p.m.
Problem Solving

- Problem solving (qualitative and quantitative) is one of the primary teaching goals, teaching tools, and evaluation techniques of physics courses.

- There is no standard way to evaluate problem solving that is valid, reliable, and easy to use.
  - Student interviews are time consuming & difficult
  - Existing rubrics are time consuming & difficult

- Need an assessment instrument for both research and instruction.

- Must consider issues of validity and reliability
  - **Validity** is the degree to which the score interpretation is supported by empirical evidence & theoretical backing.
  - **Reliability** is the stability of scores across multiple raters.
Project Goals

- Develop a robust instrument to assess students’ written solutions to physics problems, and determine reliability and validity.

- The instrument should be general
  - not specific to instructor practices or techniques
  - applicable to a range of problem topics and types

- Develop materials for appropriate use and training.

Not the most precise evaluation of problem solving
….looking for a ruler, not an electron microscope!
### Instrument at a glance (Rubric)

**CATEGORY: (based on literature)**

<table>
<thead>
<tr>
<th>Useful Description</th>
<th>Physics Approach</th>
<th>Specific Application</th>
<th>Math Procedures</th>
<th>Logical Progression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**SCORE**

<table>
<thead>
<tr>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
<th>NA (P)</th>
<th>NA (S)</th>
</tr>
</thead>
</table>

**Want**

- *Minimum number of categories that include relevant aspects of problem solving*
- *Minimum number of scores that give enough information to improve instruction*
- *Minimum training to use*

Note: 4 of the 5 categories are qualitative

Initial Version
## Rubric Scores (in general)

<table>
<thead>
<tr>
<th>Rubric Score</th>
<th>Complete &amp; appropriate</th>
<th>Minor omissions or errors</th>
<th>Parts missing and/or contain errors</th>
<th>Most missing and/or contain errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### NA Prob
- **0**: All incorrect or all missing
- **Not necessary for this problem**

### NA Solver
- **Not necessary for this solver**
A block of mass $m = 2 \text{ kg}$ slides down a frictionless ramp of height $h = 2 \text{ m}$. Use conservation of energy to determine the speed of the block at the bottom of the ramp.

**Physics Approach:**
physics concept or principle stated in problem

**Useful Description:**
visual & symbolic representation given
Pilot Study Description

- Eight experienced graduate student teaching assistants used the initial rubric to score students’ written solutions to final exam problems.

- Four volunteers scored mechanics problem solutions & four scored E&M solutions.

- After 8 solutions were scored, training consisted of example scores and rationale for the first 3 solutions. Then 5 solutions were re-scored, and 5 new solutions were scored.

- They provided written feedback on the rubric categories and scoring process.
**All Training in Writing: Example**

<table>
<thead>
<tr>
<th>CATEGORY</th>
<th>SCORE</th>
<th>RATIONALE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics Approach</td>
<td>4</td>
<td>Kinematics is appropriate before spring stretch; conservation of energy approach is explicitly stated</td>
</tr>
<tr>
<td>Useful Description</td>
<td>1</td>
<td>missing variable definitions; used “h” and “x” w/multiple values; picture is missing variable labels and height/stretch for part b)</td>
</tr>
<tr>
<td>Specific App. of Physics</td>
<td>2</td>
<td>Does not identify “initial” and “final” energy terms; part b) is missing a mgh term; used incorrect stretch value in part b)</td>
</tr>
<tr>
<td>Mathematical Procedures</td>
<td>2</td>
<td>Important algebraic mistakes when solving for k (did not need to take square root and incorrectly drops root from k)</td>
</tr>
<tr>
<td>Logical Organization</td>
<td>2</td>
<td>Should have checked units for k equation in part a) – might have caught inconsistencies;</td>
</tr>
</tbody>
</table>
## Inter-rater Agreement

<table>
<thead>
<tr>
<th></th>
<th>BEFORE TRAINING</th>
<th>AFTER TRAINING</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Perfect Agreement</td>
<td>Agreement Within One</td>
</tr>
<tr>
<td>Useful Description</td>
<td>38%</td>
<td>75%</td>
</tr>
<tr>
<td>Physics Approach</td>
<td>37%</td>
<td>82%</td>
</tr>
<tr>
<td>Specific Application</td>
<td>45%</td>
<td>95%</td>
</tr>
<tr>
<td>Math Procedures</td>
<td>20%</td>
<td>63%</td>
</tr>
<tr>
<td>Logical Progression</td>
<td>28%</td>
<td>70%</td>
</tr>
<tr>
<td>OVERALL</td>
<td>34%</td>
<td>77%</td>
</tr>
<tr>
<td>Weighted kappa</td>
<td>0.27±0.03</td>
<td>0.42±0.03</td>
</tr>
</tbody>
</table>

*Fair agreement*  
*Moderate agreement*
Findings

- NA categories and the score zero were largely ignored, even after training.
  - “[the training] Would be more helpful if it covered the 0-4 range for each category…No example of NA(P) means I still don’t know how/if to apply it.”

- Graduate student raters were influenced by their traditional grading experiences.
  - “I don't think credit should be given for a clear, focused, consistent solution with correct math that uses a totally wrong physics approach”

- The rubric works best for problems without multiple parts.
  - “[difficult] Giving one value for the score when there were different parts to the problem.”
Rubric Revisions

- The wording was made more parallel in every category.

- 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.

- Logical organization was renamed logical progression.
Next Steps

- **Expand 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


Rubric Categories (based on research literature)

- **Useful Description**
  - organize information from the problem statement symbolically, visually, and/or in writing.

- **Physics Approach**
  - select appropriate physics concepts and principles to use

- **Specific Application of Physics**
  - apply physics approach to the specific conditions in problem

- **Mathematical Procedures**
  - follow appropriate & correct math rules/procedures

- **Logical Progression**
  - (overall) solution progresses logically; it is coherent, focused toward a goal, and consistent

**Note:** 4 of the 5 categories are qualitative
Range of detail in solutions

Useful Description: unnecessary for this solver NA(S)