UNIVERSITY OF MINNESOTA

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.***

I am Jennifer from the University of Minnesota PER group.

This will be very abbreviated…if you want more detailed information, visit my poster today.

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.

•Problem solving is an important part of learning physics.

•Despite this, there is no standard way to evaluate problem solving that is VALID, RELIABLE, and EASY TO USE. Students interviews will give you a lot of information, but (as some of you in this room know first-hand) they are time consuming and difficult to administer and analyze. Existing instruments that I'm aware of (from UMN and other institutions) are complex and require training, or they haven't been extensively tested.

•Such an instrument will benefit both the research community and instructors.

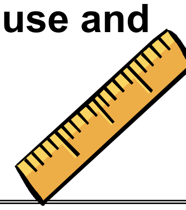
•Anytime you're doing instrument development, you need to consider issues of validity and reliability.

•Modern validity theory has moved away from different "types" of "kinds" of validity to focus on different kinds of evidence from data and the research literature to support interpretations of scores on the instrument. Reliability in this context refers to the consistency of different people using the instrument.

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!



- The goal of this project is to develop such an instrument to assess WRITTEN solutions to physics problems

- The instrument should be independent of how a student was taught to solve problems (not biased to a particular strategy) and applicable to a variety of physics topics (mechanics, E&M) and problem types (both quantitative and qualitative problems).

- This research will also develop materials for training.

- The punch line is….we're not interested in the most precise measure of problem solving; we're looking for a ruler, not an electron microscope!

Instrument at a glance (Rubric)

SCORE

CATEGORY: (based on literature)	4	3	2	1	0	NA (P)	NA (S)
Useful Description							
Physics Approach							
Specific Application							
Math Procedures							
Logical Progression							

Initial Version

Note: 4 of the 5 categories are qualitative

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*

•This is the instrument we're working on, at a glance. It takes the form of a rubric (which is a table or grid). It identifies five problem-solving sub-skill categories and defines performance levels for each category by a score and the criteria met to attain that score (empty boxes in this picture). If you want to see the rubric in its entirety, visit my poster.

•The categories are based on the research literature from cognitive psychology and physics education, and are intended to be somewhat independent. The idea is that a student receives a separate score for each category, and this will give a more detailed description of a student's strengths and weaknesses in order to direct instruction.

•For example, a student could identify appropriate physics principles and concepts to apply to the problem (physics approach) but have difficulty applying it to the specific conditions in the problem (specific application of physics). Or a student could have the physics correct, but get hung up on the math procedures.

•For simplicity, we're looking for the MINIMUM number of categories that will still include relevant aspects of problem solving, and the MINIMUM number of scores that will still give enough information to improve instruction. And we don't want it to require a lot of training.

Rubric Scores (in general)

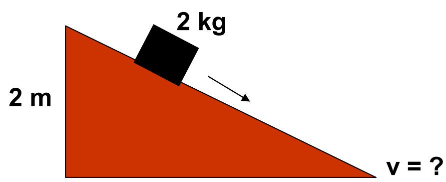
4	3	2	1
Complete & appropriate	Minor omissions or errors	Parts missing and/or contain errors	Most missing and/or contain errors

0	NA Prob	NA Solver
All incorrect or all missing	Not necessary for this problem	Not necessary for this solver

•In General, the rubric scores range from complete and appropriate (4) to minor and more serious errors (3-1) and all incorrect/missing.

•The next two categories represent NA or "not applicable". The NA Problem category means that a particular skill was not necessary for the problem, and I'll show you an example of this on the next slide. The NA Solver category means that based on the overall solution, it was not necessary for the solver to explicitly write down that step in the problem.

Example of NA (Problem)



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

A block of mass *m* = 2 kg slides down a frictionless ramp of height *h* = 2 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

•This is an example of a problem in which the visual description and symbolic notation has been provided for the solver in the problem statement.

•Also, the appropriate physics principle has already been identified for the solver (“use conservation of energy”).

•In scoring this problem with the rubric, the Useful Description and Physics Approach categories would both receive a score of NA-Problem.

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.

•So, how have we tested it?

•Last fall I recruited eight experienced graduate students (3rd year or higher) to use the rubric to score student solutions to archived final exams.

•Half scored a mechanics problem, and half scored E&M.

•The procedure was to score 8 solutions, receive written training materials and documentation for the first 3 solutions (which consisted of example scores and rationale) and then to re-score 5 solutions and score 5 new solutions.

•The graduate students also provided written feedback on the rubric. I asked them which category was most difficult to score and why, and suggestions for improvement to the instrument.

All Training in Writing: Example

CATEGORY

SCORE

RATIONALE

Training includes the actual student solution

Student # 2	Score	Notes
Physics Approach	**4**	Kinematics is appropriate before spring stretch; conservation of energy approach is explicitly stated
Useful Description	**1**	missing variable definitions; used "h" and "x" w/multiple values; picture is missing variable labels and height/stretch for part b)
Specific App. of Physics	**2**	Does not identify "initial" and "final" energy terms; part b) is missing a mgh term; used incorrect stretch value in part b)
Mathematical Procedures	**2**	Important algebraic mistakes when solving for k (did not need to take square root and incorrectly drops root from k)
Logical Organization	**2**	Should have checked units for k equation in part a) – might have caught inconsistencies;

This is what the written "training" looked like. Graduate students received example scores and rationale for three problems so they could compare these scores to their own and match them up to the actual student solutions.

Inter-rater Agreement

	BEFORE TRAINING		AFTER TRAINING	
	Perfect Agreement	Agreement Within One	Perfect Agreement	Agreement Within One
Useful Description	38%	75%	38%	80%
Physics Approach	37%	82%	47%	90%
Specific Application	45%	95%	48%	93%
Math Procedures	20%	63%	39%	76%
Logical Progression	28%	70%	50%	88%
OVERALL	34%	77%	44%	85%
Weighted kappa	*0.27±0.03*		*0.42±0.03*	

Fair agreement → *Moderate agreement*

Slide 9 of 12 Jennifer Docktor - Robust Assessment Instrument for Student Problem Solving 7/22/2008

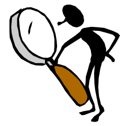
•What did we learn?

•We compared the raters' scores with those of an expert rater (namely, me) and measured both the perfect agreement and agreement within one score…For example, if I gave a score of 3 and the rater gave a score of 4, this is agreement within one.

•It was actually pretty good before training (77%) and improved slightly to 85% after training. The categories that improved the most were math procedures and logical progression.

•According to a more statistical measure of agreement called quadratic weighted kappa, there was fair agreement above chance before training that improved to moderate agreement after training.

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."*

•What else did we find?

•The NA categories were confusing, and were not used. Raters would like to see more examples of when these scores are applicable. The zero score was also rarely used; even if something was entirely incorrect, grad students gave a score of 1 for the attempt.

•Also, graduate students had difficulty treating the categories independently. For example, they were unwilling to give points for math and logic if the physics approach was incorrect.

•The mechanics problem included parts a) and b), and at least two of the four raters on this problem expressed they were unsure whether to give scores for each part separately, or give an overall score. For this reason, multi-part problems are difficult to score with the instrument.

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

Revisions to the instrument as a result of the pilot.

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).

•What are the next steps?

•We need to expand the training materials to include a greater range of examples that span the scores in each category, especially NA scores. In the documentation we need to specifically address the rubric's purpose, to distinguish it from traditional grading practices.

•Then we need to re-test revised rubric + training materials and re-assess reliability.

•After we have the reliability pinned down, we need to think about validity measures.

Poster: PST4-25   5-6 p.m.

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

docktor@physics.umn.edu

References

P. Heller, R. Keith, and S. Anderson, "Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving," Am. J. Phys., 60(7), 627-636 (1992).

J.M. Blue, *Sex differences in physics learning and evaluations in an introductory course.* Unpublished doctoral dissertation, University of Minnesota, Twin Cities (1997).

T. Foster, *The development of students' problem-solving skills from instruction emphasizing qualitative problem-solving.* Unpublished doctoral dissertation, University of Minnesota, Twin Cities (2000).

J.H. Larkin, J. McDermott, D.P. Simon, and H.A. Simon, "Expert and novice performance in solving physics problems," Science 208 (4450), 1335-1342.

F. Reif and J.I. Heller, "Knowledge structure and problem solving in physics," Educational Psychologist, 17(2), 102-127 (1982).

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, *Standards for educational and psychological testing* (Washington, DC: American Educational Research Association, 1999).

P.A. Moss, "Shifting conceptions of validity in educational measurement: Implications for performance assessment," Review of Educational Research 62(3), 229-258 (1992).

J. Cohen, "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit," Psychological Bulletin 70(4), 213-220 (1968).

You can download the rubric and documentation at the UMN PER website (by clicking on my name) or send me an e-mail.

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

(ADDITIONAL SLIDE – was cut out)

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

Reflect stages in the physics problem-solving process

Intended to be somewhat independent.

Range of detail in solutions

$\overset{+}{\longrightarrow}_{F_e}$ $qE = F_e = ma$

$E = \frac{ma}{q}$ $\quad v_f^2 = \cancel{v_i^2}^{\,0} + 2a\Delta x$

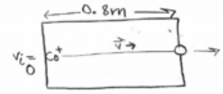
$a = \frac{v_f^2}{2\Delta x}$

$E = \frac{m v_f^2}{q\, 2\Delta x} = \frac{[kg]\cdot[m]^{\cancel{2}}}{[s]^2[C]\cancel{[m]}} = \frac{N}{C}$ $\quad \frac{V}{m} = \frac{N}{C}$ $\quad N = \frac{kg\, m}{s^2}$

units ok ✓

$\boxed{E = 1162\ \frac{N}{C}}$ "

***Useful Description*: unnecessary for this solver NA(S)**



$s = 0.8$ m
$v_f = 8\times10^4$ m/s
$m_{co} = 28$ g/mol

Question: Calculate the direction and magnitude of the electric field needed so that CO^+ ions created at rest at one end will have a speed of 8×10^4 m/s when they exit the other side

Approach: Use conservation of energy

system: Co^+ particle

initial time: right as CO^+ enters the box

final time: right as Co^+ leaves the box

$E_i = 0 \quad v_i = 0$

$E_f = \frac{1}{2}mv^2$, E_{in} = electric potential energy, $E_{out} = 0$

Electric potential energy = $\Delta V q = -q\int \vec{E}\cdot ds$

because the electric field is constant

$PE_e = -qE\int ds$

$PE_e = -qES$ ← just want magnitude so can leave negative sign off

$E_f - E_i = E_{in} - E_{out}$ $\qquad$ check units

$\frac{1}{2}mv^2 = qES$ $\qquad$ $\frac{1}{2}mv^2 \rightarrow$ energy units

$E = \frac{\frac{1}{2}mv^2}{qS}$ $\qquad$ $\frac{E\cdot S,\ q}{V\cdot C}$ = energy units

" units ok ✓

$|E| = \frac{\frac{1}{2}mv^2}{qS} =$ $\qquad$ $J = N\cdot m$

$CO - e^- = CO^+$

$.028\text{ kg} - 9.11\times10^{-31}\text{ kg} = 0.028\text{ kg}$

$|E| = \frac{\frac{1}{2}(0.028\text{ kg})(8\times10^4\text{ m/s})^2}{(1.602\times10^{-19}\text{ C})(0.8\text{ m})}$

$\boxed{|E| = 6.99\times10^{26}\text{ N/C}}$

(ADDITIONAL SLIDE)