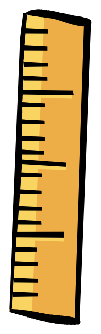
Robust Assessment Instrument for Student Problem Solving

Jennifer Docktor, Kenneth Heller

University of Minnesota

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

INTRODUCTION



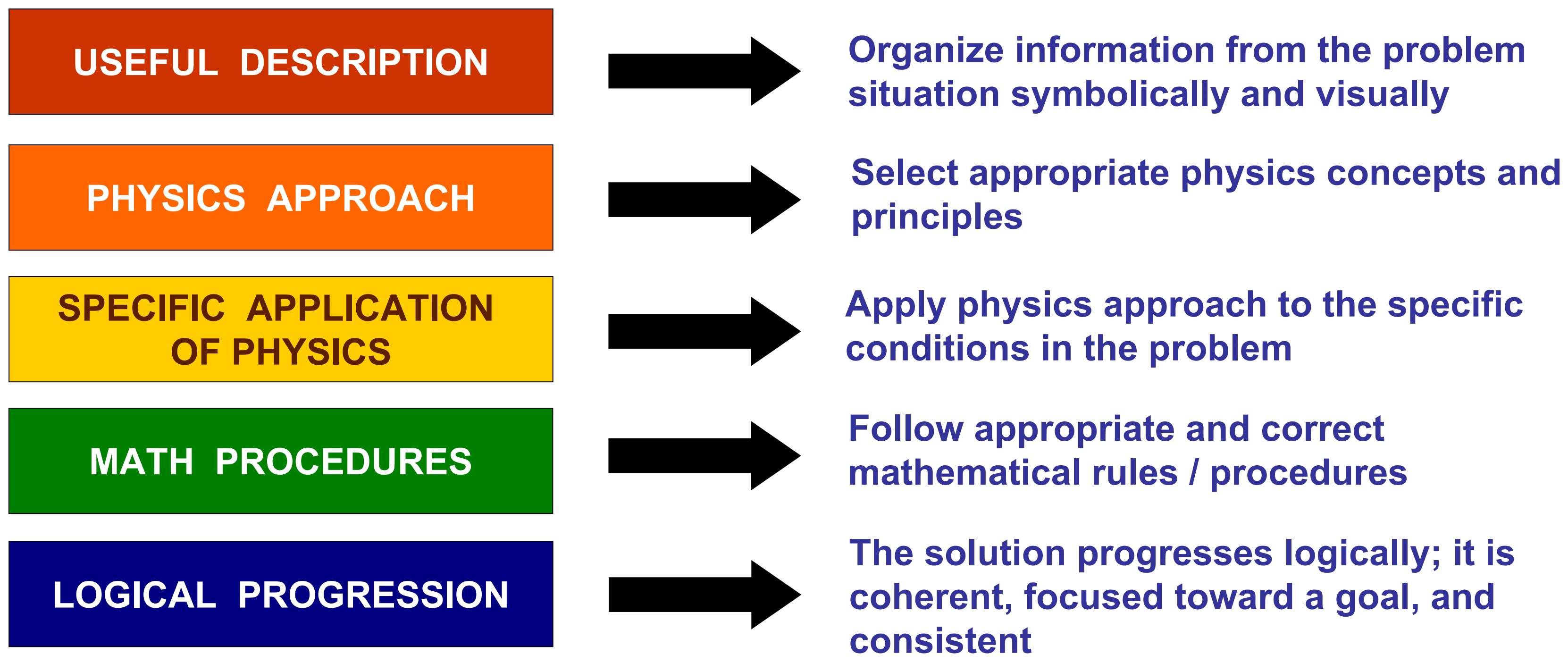
Problem solving skills (qualitative and quantitative) are a primary tool used in most physics instruction. Despite this importance, a reliable, valid, and easy to use quantitative measure of physics problem solving does not exist. Although scoring tools have been used in past problem solving research at the University of Minnesota[1-3] 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.[4-8]

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.[9-12]



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

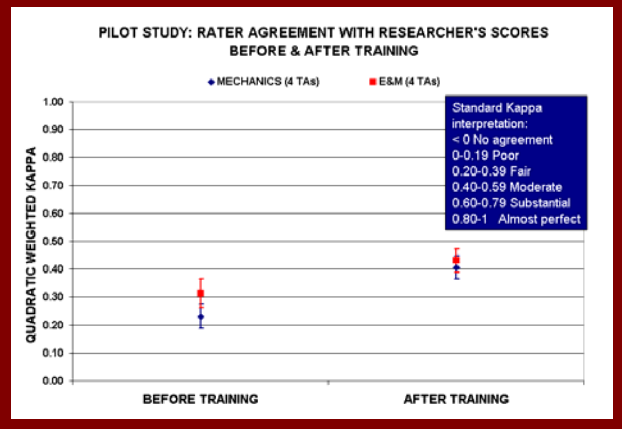
LOGICAL PROGRESSION → The solution progresses logically; it is coherent, focused toward a goal, and consistent

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

	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 Measure of Agreement[13]:

$$\kappa_w = \frac{\sum w_{ij} f_{oij} - \sum w_{ij} f_{Eij}}{(w_{\max})N - \sum w_{ij} f_{Eij}}$$

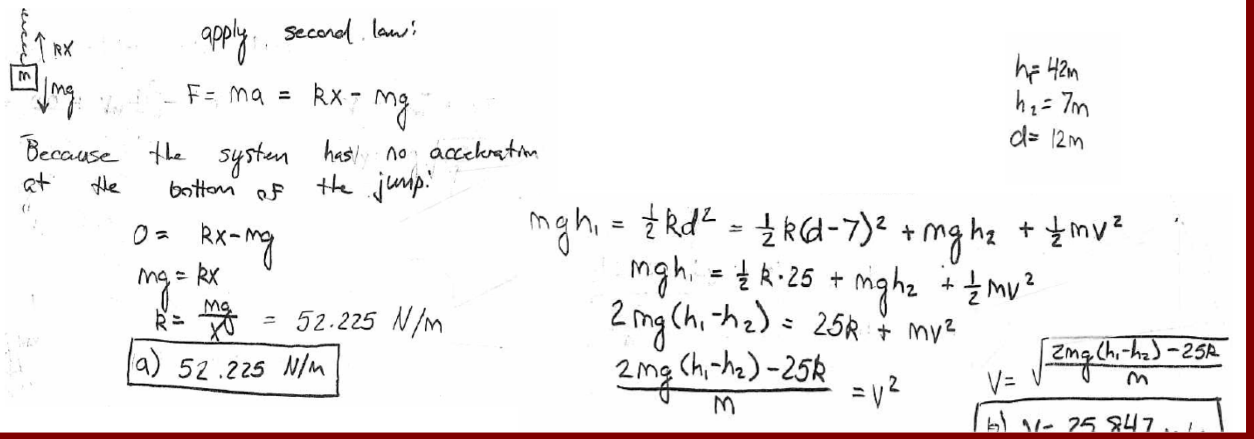
f_o: observed frequencies

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 convince the dean that your plan is safe for a person of his mass, 70 kg. Your plan has the dean stepping off a platform and being in free fall for 30 m before the cord begins to stretch. Determine the spring constant of the bungee cord so that it stretches only 12m, which will just keep the dean out of the water. Find the dean's speed 7m above the water.



apply second law:

$F = ma = RX - mg$

Because the system has no acceleration at the bottom of the jump.

$0 = RX - mg$

$mg = RX$

$R = \frac{mg}{X} = 52.225\ N/m$

$\boxed{a)\ 52.225\ N/m}$

$h_1 = 42m$
$h_2 = 7m$
$d = 12m$

$mgh_1 = \frac{1}{2}Rd^2 = \frac{1}{2}R(d-7)^2 + mgh_2 + \frac{1}{2}mv^2$

$mgh_1 = \frac{1}{2}R \cdot 25 + mgh_2 + \frac{1}{2}mv^2$

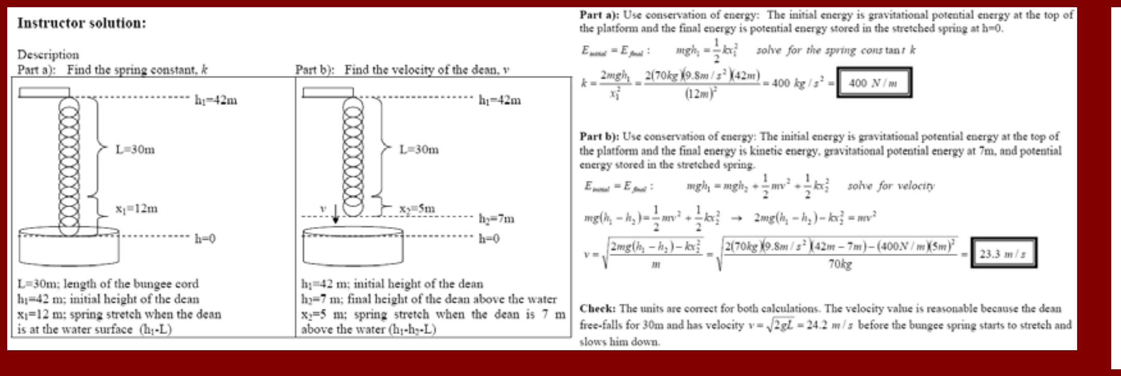
$2mg(h_1 - h_2) = 25R + mv^2$

$\frac{2mg(h_1-h_2) - 25R}{m} = v^2$

$V = \sqrt{\frac{2mg(h_1-h_2)-25R}{m}}$

b) V= 25.847

TRAINING



Instructor solution:

Description

Part a): Find the spring constant, k

Part b): Find the velocity of the dean, v

L=30m; length of the bungee cord
h1=42 m; initial height of the dean
x1=12 m; spring stretch when the dean is at the water surface (h1-L)

h1=42 m; initial height of the dean
h2=7 m; final height of the dean above the water
x2=5 m; spring stretch when the dean is 7 m above the water (h1-h2-L)

Part a): Use conservation of energy: The initial energy is gravitational potential energy at the top of the platform and the final energy is potential energy stored in the stretched spring at h=0.

$E_{initial} = E_{final}: \quad mgh_1 = \frac{1}{2}kx_1^2$ solve for the spring constant k

$k = \frac{2mgh_1}{x_1^2} = \frac{2(70kg)(9.8m/s^2)(42m)}{(12m)^2} = 400\ kg/s^2 = \boxed{400\ N/m}$

Part b): Use conservation of energy: The initial energy is gravitational potential energy at the top of the platform and the final energy is kinetic energy, gravitational potential energy at 7m, and potential energy stored in the stretched spring.

$E_{initial} = E_{final}: \quad mgh_1 = mgh_2 + \frac{1}{2}mv^2 + \frac{1}{2}kx_2^2$ solve for velocity

$mg(h_1 - h_2) = \frac{1}{2}mv^2 + \frac{1}{2}kx_2^2 \rightarrow 2mg(h_1 - h_2) - kx_2^2 = mv^2$

$v = \sqrt{\frac{2mg(h_1-h_2) - kx_2^2}{m}} = \sqrt{\frac{2(70kg)(9.8m/s^2)(42m-7m) - (400N/m)(5m)^2}{70kg}} = \boxed{23.3\ m/s}$

Check: The units are correct for both calculations. The velocity value is reasonable because the dean free-falls for 30m and has velocity $v = \sqrt{2gL} = 24.2\ m/s$ before the bungee spring starts to stretch and slows him down.

Student # 1	Score	Notes
Physics Approach	1	Newton's second law is inappropriate during spring stretch; missing energy conservation for part a); approach in b) is unclear
Useful Description	NA(S)	Visualization is unnecessary for this solver; Free-body diagram assumes = forces; defined variables for part b) but not a);
Specific App. of Physics	2	Incorrectly assumes acceleration is zero at bottom of jump; does not identify "initial" and "final" energy terms
Mathematical Procedures	3	Missing substitution of numerical values during calculations (except d-7); makes a calculation error when finding k in part a)
Logical Organization	2	Parts of the solution are unclear due to implicit reasoning; velocity value is greater than free fall after 30 m;

Grant DUE- 0715615

doctkor@physics.umn.edu

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

	5	4	3	2	1	0	NA (Problem)	NA (Solver)
Useful Description	The description is useful, appropriate, and complete.	The description is useful but contains minor omissions or errors.	Parts of the description are not useful, missing, and/or contain errors.	Most of the description is not useful, missing, and/or contains errors.	The entire description is not useful and/or contains errors.	The solution does not include a description and it is necessary for this problem /solver.	A description is not necessary for this problem. (i.e., it is given in the problem statement)	A description is not necessary for this solver.
Physics Approach	The physics approach is appropriate and complete.	The physics approach contains minor omissions or errors.	Some concepts and principles of the physics approach are missing and/or inappropriate.	Most of the physics approach is missing and/or inappropriate.	All of the chosen concepts and principles are inappropriate.	The solution does not indicate an approach, and it is necessary for this problem/ solver.	A physics approach is not necessary for this problem. (i.e., it is given in the problem statement)	An explicit physics approach is not necessary for this solver.
Specific App. of Physics	The specific application of physics is appropriate and complete.	The specific application of physics contains minor omissions or errors.	Parts of the specific application of physics are missing and/or contain errors.	Most of the specific application of physics is missing and/or contains errors.	All of the application of physics is inappropriate and/or contains errors.	The solution does not indicate an application of physics and it is necessary.	Specific application of physics is not necessary for this problem.	Specific application of physics is not necessary for this solver.
Mathematical Procedures	The mathematical procedures are appropriate and complete.	Appropriate mathematical procedures are used with minor omissions or errors.	Parts of the mathematical procedures are missing and/or contain errors.	Most of the mathematical procedures are missing and/or contain errors.	All mathematical procedures are inappropriate and/or contain errors.	There is no evidence of mathematical procedures, and they are necessary.	Mathematical procedures are not necessary for this problem or are very simple.	Mathematical procedures are not necessary for this solver.
Logical Progression	The entire problem solution is clear, focused, and logically connected.	The solution is clear and focused with minor inconsistencies.	Parts of the solution are unclear, unfocused, and/or inconsistent.	Most of the solution parts are unclear, unfocused, and/or inconsistent.	The entire solution unclear, unfocused, and/or inconsistent.	There is no evidence of logical progression, and it is necessary.	Logical progression is not necessary for this problem. (i.e., one-step)	Logical progression is not necessary for this solver.

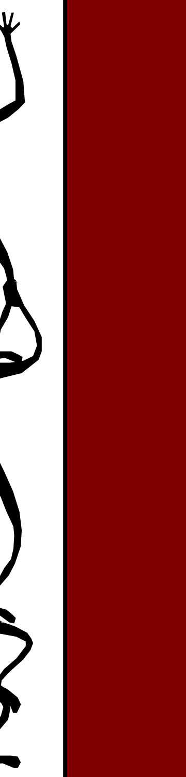
FINDINGS

- 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 rubric works best for problems that do not have multiple parts.
  - *"[difficult] Giving one value for the score when there were different parts to the problem." [TA #2]*
- Written comments indicate the graduate student raters were influenced by their traditional grading experiences. They expressed concerns about giving "too much credit" for math and logical progression if the physics was inappropriate, and had difficulty distinguishing categories.
  - *"[The student] didn't do any math that was wrong, but it seems like too many points for such simple math"[TA#8]*
  - *"Specific application of physics was most difficult. I find this difficult to untangle from physics approach. Also, how should I score it when the approach is wrong?" [TA#1]*

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



[1] 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).

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

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

[4] 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).

[5] S. Messick, "Validity of psychological assessment," American Psychologist 50(9), 741-749 (1995).

[6] M.T. Kane, "An argument-based approach to validity," Psychological Bulletin 112(3), 527-535 (1992).

[7] M.T. Kane, "Current concerns in validity theory," Journal of Educational Measurement 38(4), 319-342 (2001).

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

[9] 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.

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

[11] M.T.H. Chi, P. Feltovich, and R. Glaser, "Categorization and representation of physics problems by experts and novices," Cognitive Science 5, 121-152 (1981).

[12] A.H. Schoenfeld, *Mathematical problem solving* (Orlando, FL: Academic Press, Inc., 1985).

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