Robust Assessment Instrument for Student Problem Solving

Jennifer L. Docktor

Ken Heller, Leon Hsu

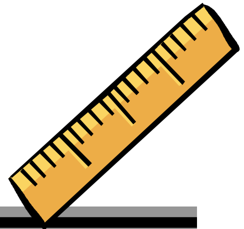
UNIVERSITY OF MINNESOTA

Physics Education Research & Development Group

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

DUE-0715615

Problem Solving Measure



- Problem solving is both an important mechanism and outcome of learning.
  - This is certainly true in physics
  - Unfortunately, there is no standard way to easily measure problem solving so that student progress can be assessed.
  - Our *goal* is to develop an easy-to-use, robust instrument to assess students' <u>written</u> solutions to physics problems, and obtain evidence for reliability, validity, and utility of scores.
- The instrument should be *general*
  - not specific to instructor practices or techniques
  - applicable to a range of problem topics and types

Reliability, Validity, & Utility

- ***Reliability*** **– score agreement**
- ***Validity*** **evidence from multiple sources**
  - **Content (relevant & representative)**
  - **Response processes**
  - **Internal & external structure**
  - **Generalizability**
  - **Consequences of testing**
- ***Utility*** **- usefulness of scores**

Assessment construction must address these concepts

AERA, APA, NCME (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, 50(9), 741-749.

Instrument at a Glance (Rubric)

SCORE ←

CATEGORY:	5	4	3	2	1	0	NA (Prob)	NA (Slvr)
Useful Description								
Physics Approach								
Specific Application								
Math Procedures								
Logical Progression								

Separate score for each category – indicates strengths & weaknesses

Minimize the number of categories & scores

Research Basis for Rubric

Representative research literature

- Categories come from:
  - Problem solving processes
    - Pólya (1945), Newell & Simon (1972), Reif & J. Heller (1982, 1984), Schoenfeld (1985), Van Heuvelen (1991)
  - Expert-novice characteristics
    - Chi, Feltovich, & Glaser (1981), Larkin (1979), Larkin, McDermott, Simon, & Simon (1980), Hardiman, Dufresne, & Mestre (1989), Singh (2002, 2008)
  - Previous work at Minnesota
    - P. Heller, Keith, & Anderson (1992), Blue (1997), Foster (2000)
- Instrument construction:
  - Validity, Reliability, Utility
    - AERA, APA, NCME (1999), Kane (1992), Messick (1995), Moss (2007), Cohen (1968)
  - Rubrics
    - Arter & McTighe (2001), Montgomery (2002)

Rubric Scores (in general)

5	4	3	2	1	0
Complete & appro-priate	Minor omission or errors	Parts missing and/or contain errors	Most missing and/or contain errors	All inappro-priate	No evidence of category

NOT APPLICABLE (NA):

NA - Problem	NA - Solver
Not necessary for this problem *(i.e. visualization or physics principles given)*	Not necessary for this solver *(i.e. able to solve without explicit statement)*

Overview of Study

1. Drafting the rubric
2. Preliminary tests with two raters (final exams and instructor solutions)
3. Training exercise with graduate students
4. Analysis of tests from an introductory mechanics course
5. Student problem-solving interviews (in progress)

Initial Training Exercise

Results - preliminary rubric use

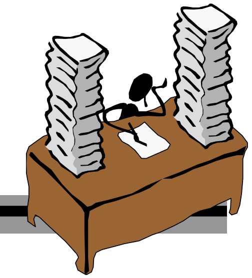
8 physics graduate students with TA experience

- **Score agreement improved significantly with minimal training**
  - ***Weighted kappa 0.27±0.03 Fair → 0.42±0.03 Moderate***
  - **Math & Logical Progression most affected**
- **Raters influenced by traditional grading experience**
  - **Unwilling to score math and logic if physics incorrect**
- **Multi-part problems more difficult to score**
- **Revisions to rubric and training based on this**
  - **Consistent language across rubric categories**
  - **More examples of NA scores & expand zero score**
  - **Distinguish physics approach & application**

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

Cohen's Weighted Kappa

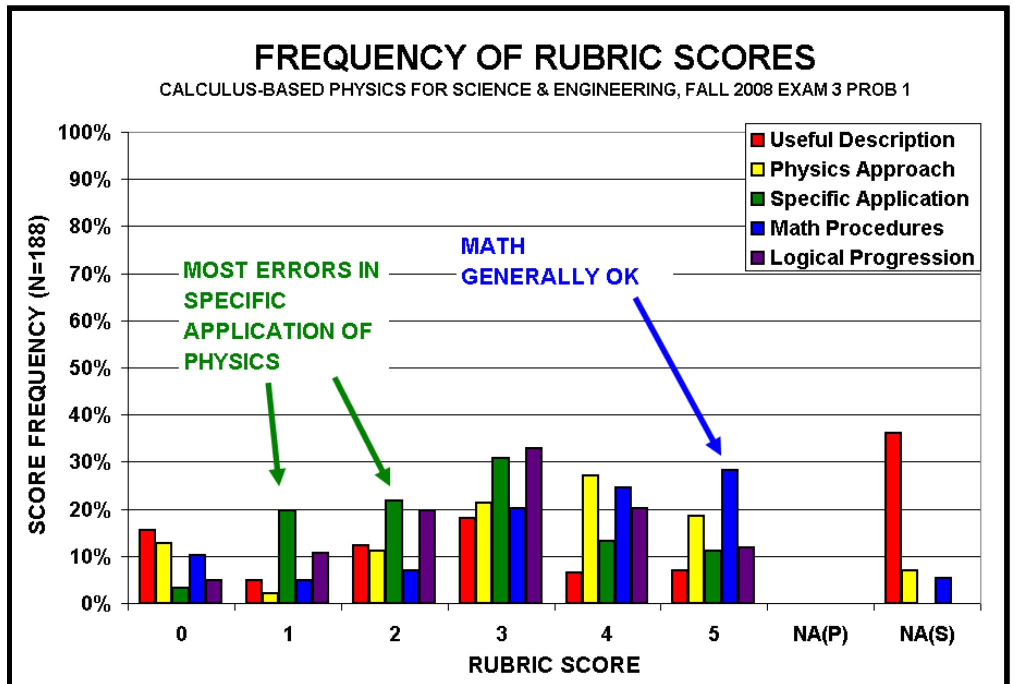
Analysis of Tests



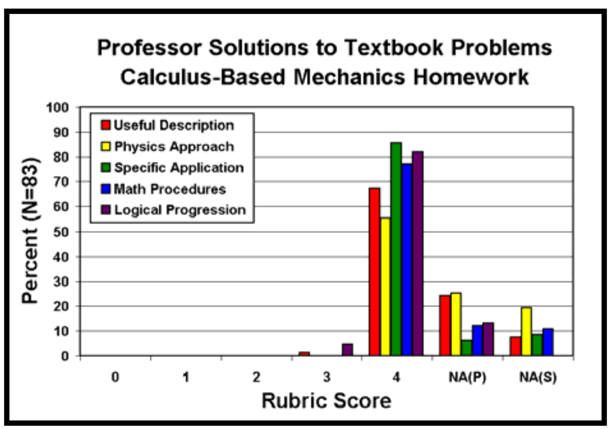
- Calculus-based introductory physics course for Science & Engineering students (mechanics)
- Test problems graded in the usual way by teaching assistants, then scored with rubric by researcher

EXAMPLE DATA

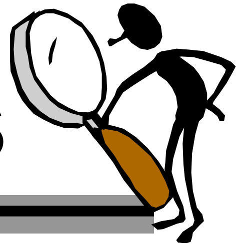
Student Solutions



Instructor Solutions



Findings from Test Analysis

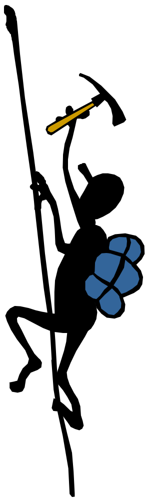


- The rubric indicates areas of student difficulty for a given problem
  - Focus instruction to coach physics, math, clear and logical reasoning processes, etc.

- The rubric responds to different problem features
  - Can make score interpretation difficult
    - usefulness of visualization
    - prompts & cues
    - numeric vs. symbolic question
  - Modify problems to elicit / practice processes

Summary

- A rubric is being developed from descriptions of problem solving process, expert-novice research studies, and past studies at UMN
  - Focus on written solutions to physics problems
- Training revised to improve score agreement
- Rubric provides useful information that can be used for research & instruction
- Rubric works for standard range of physics topics in an introductory mechanics course
  - There are some problem characteristics that make score interpretation difficult (prompts & cues)
- Interviews will provide information about response processes



For more information:

docktor@physics.umn.edu

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

DUE-0715615

Additional Slides

References - Categories

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).

P. Hardiman, R. Dufresne, and J. Mestre, "The relation between problem categorization and problem solving among experts and novices," *Memory & Cognition*, *17*(5), 627-638 (1989).

J. Heller, and F. Reif, "Prescribing effective human problem solving processes: Problem description in physics," *Cognition and Instruction*, *1*(2), 177-216 (1984).

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. Larkin, "Processing information for effective problem solving," *Engineering Education*, *70*(3), 285-288, (1979).

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, (1980).

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

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

C. Singh, "When physical intuition fails," *Am. J. Phys.*, *70*(11), 1103-1109 (2002).

Van Heuvelen, "Overview, case study physics," *Am. J. Phys.*, *59*(10), 898-907 (1991)

References – Instrument Construction

AERA, APA, and NCME, *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association (1999).

J. Arter and J. McTighe, *Scoring rubrics in the classroom*. Thousand Oaks, CA: Corwin Press, Inc. (2001).

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

M.T. Kane, “An argument-based approach to validity,” *Psychological Bulletin*, *112*(3), 527-535 (1992).

S. Messick, “Validity of psychological assessment,” *American Psychologist*, *50*(9), 741-749 (1995).

K. Montgomery, “Authentic tasks and rubrics: Going beyond traditional assessments in college teaching,” *College Teaching*, *50*(1), 34-39 (2002).

P.A. Moss, “Reconstructing validity,” *Educational Researcher*, *36*(8), 470-476 (2007).

Weighted Kappa

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

$$\sigma_{k\,w} = \sqrt{\frac{N \sum w^2{}_{ij} f_{oij} - \left(\sum w_{ij} f_{oij}\right)^2}{N\left(w_{\max} N - \sum w_{ij} f_{Eij}\right)^2}}$$

fo: observed frequencies of exact agreement (diagonal of pivot table)

fe: expected frequencies of exact agreement by chance

N: number of subjects rated

$$z = \frac{\kappa_{w1} - \kappa_{w2}}{\sqrt{\sigma^2_{\kappa_{w1}} - \sigma^2_{\kappa_{w2}}}}$$

95% confidence limit: 1.960

99% confidence limit: 2.576

99.9% confidence limit: 3.291

Kappa

$$\kappa = \frac{\sum f_o - \sum f_E}{N - \sum f_E}$$

fo: observed frequencies of exact agreement (diagonal of pivot table)

fe: expected frequencies of exact agreement by chance

N: number of subjects rated

$$\sigma_\kappa = \sqrt{\frac{N \sum f_o - \left(\sum f_o\right)^2}{N\left(N - \sum f_E\right)^2}}$$

$$z = \frac{\kappa_1 - \kappa_2}{\sqrt{\sigma_{\kappa_1}^2 - \sigma_{\kappa_2}^2}}$$

95% confidence limit: 1.960

99% confidence limit: 2.576

99.9% confidence limit: 3.291