Examining Exams

Chemistry Education Brown Bag Lunch

April 24, 2014

E. Jane Maxwell

Items in BLUE were added during the workshop to reflect the group discussion.

Examining Exams

Questions for today:

1. How can the design of an exam affect what it measures?

2. How can we tell if an exam has met its goals?

3. What can we learn from an exam once it's over?

Other questions…?

How can the design of an exam affect what it measures?

Group discussion: What factors affect students' performance on exams?

- ***Preparation***
- *Motivation*
- *Stress level*
  - *Due to current exam*
  - *Due to other demands*
- *Health (sleep, eating)*
- *Appropriate practice*
- *Expectations*
- *Maturity*
- *Clarity of questions*
- *Reading comprehension*
- *Time pressure*
- ***Time management***
- *Exam-writing strategy*
- *Study strategy*
- ***Knowledge***

Which factors do we *wish* to measure? Indicated above in **Bold**

Note that only **3** of the 14 factors are things that we wish to measure! Other factors will only cause interference in measuring the 3 of interest.

How can the **format** of an exam affect the result?

R Lukhele, D Thissen, and H Wainer. **On the Relative Value of Multiple-Choice, Constructed Response, and Examinee-Selected Items on Two Achievement Tests.** Journal of Educational Measurement, V. 31, N. 3, pp. 234-250 (1994).

1989 Chemistry AP Test (Full)

75 MCQs
(1 point each)

Part A: Long answer
1 Problem (9 points)

Parts B-D: Choice
Choose 9 of 15 long-answer problems
(9 points each)

vs. **"Shortened" versions**

75 MCQs
(1 point each)

75 MCQs
(1 point each)

2 Long Answer
(9 points each)

75 MCQs
(1 point each)

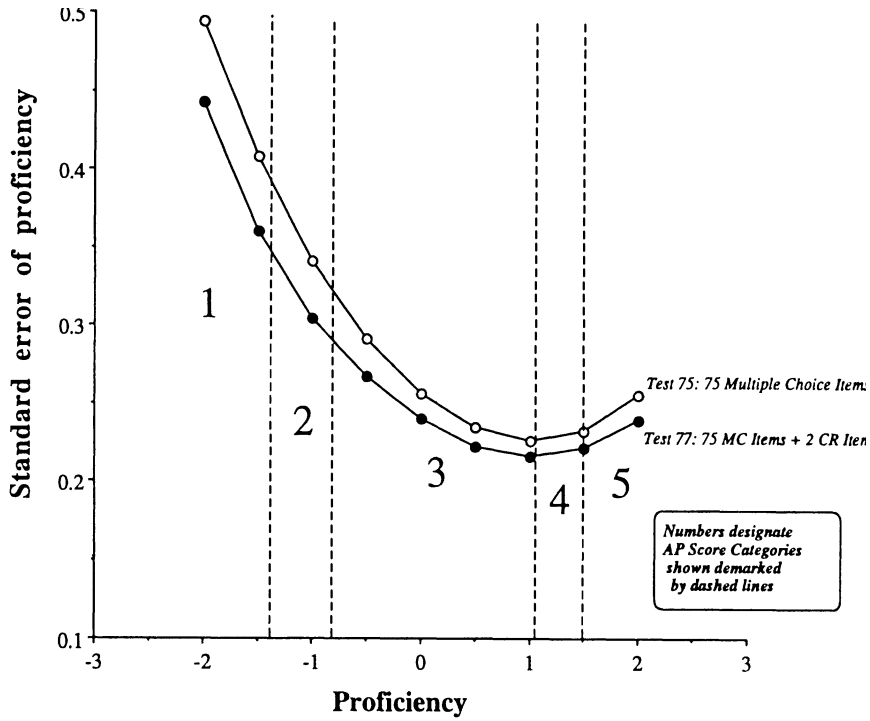
2 Long Answer
(9 points each)

4 Choice problems
Answer 4 of 7
(9 points each)

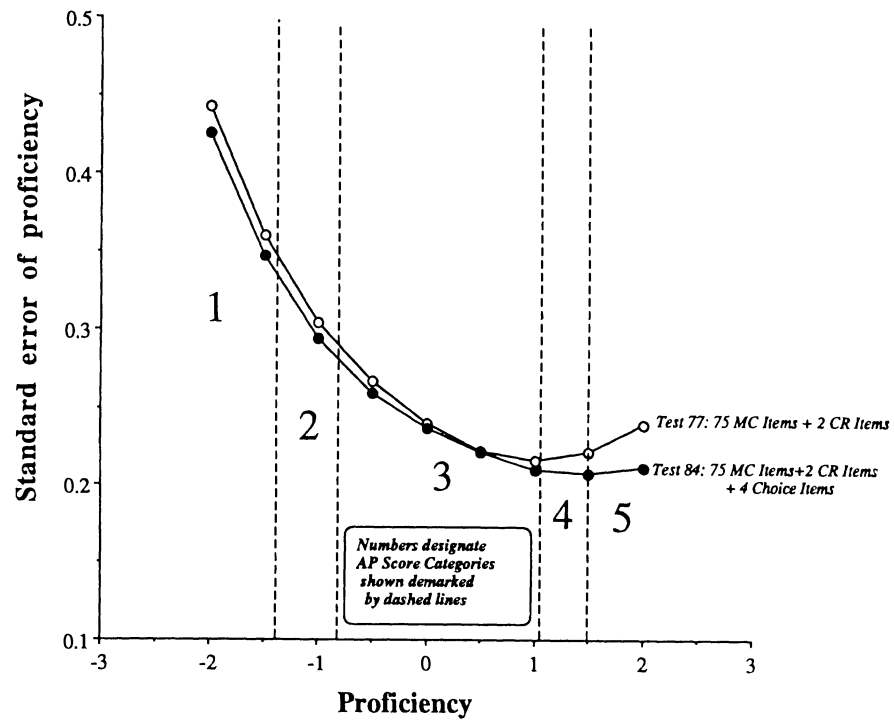
Lukhele et al. compared scores achieved by grading the entire AP exam with those acieved by grading only a subset of the exam questions ("shortened" versions).

How can the **format** of an exam affect the result?

MCQs <u>only</u> (O) *vs.*
MCQs + 2 long-answer questions (●)



MCQs + 2 long-answer questions (O) *vs.*
MCQs + 2 LA questions + 4 choice Qs (●)



Standard error: *difference between the full exam grade and a grade based on a shortened version.*

The multiple choice section was actually a very good predictor of students' overall exam score. Adding constructed-response and choice items made only small differences in grades, but added significantly to the exam length (TIME).

Lukhele et al.

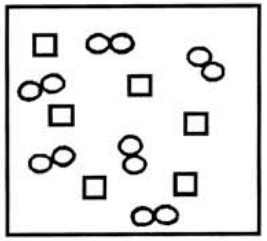
How can the type of *thinking* required affect the result?

1. Calculate the maximum weight of SO_3 that could be produced from 1.9 mol of oxygen and excess sulfur. **71%**

2. The equation for a reaction is:
   $2\ S + 3\ O_2 \rightarrow 2\ SO_3$

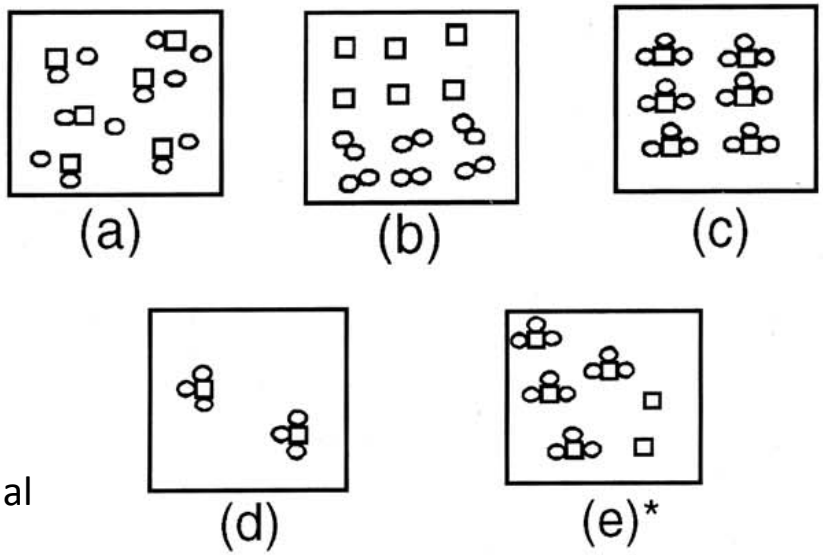
   Consider the mixture of sulfur (□) and O_2 (∞) in a closed container illustrated at right.

   Which of the following (a-e) represents the product mixture?



46%

Which of the following represents the product mixture?



Based on 546 students at Purdue University

Zoller et al. Success on algorithmic and LOCS vs. conceptual chemistry exam questions. J. Chem. Ed., 1995

How can we tell if an exam has met its goals?

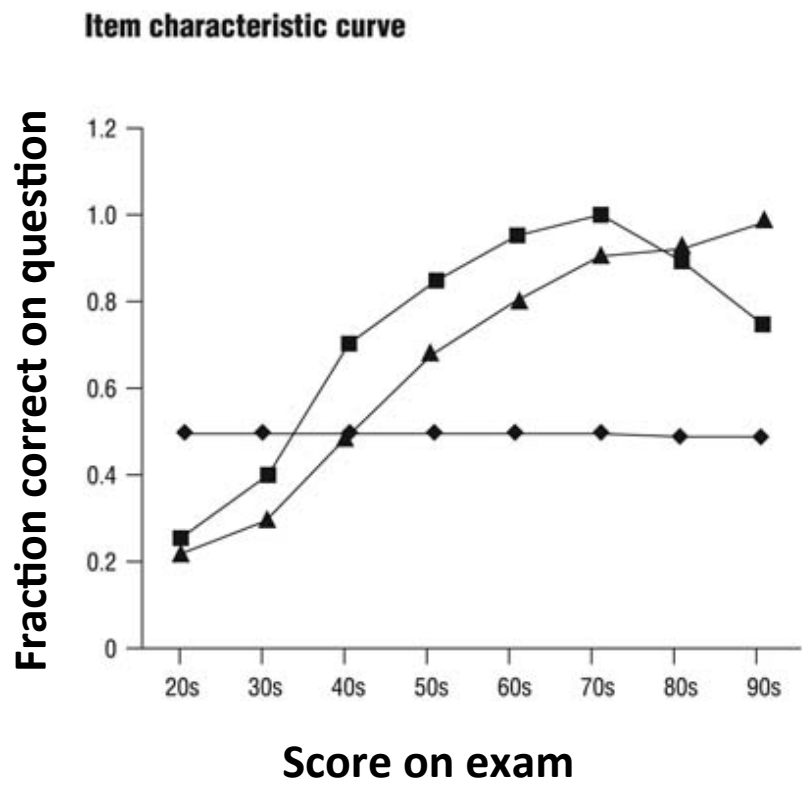
Group discussion of exam GOALS:

- *Differentiated between levels of ability*
  - *Subject mastery*
  - *Application of learning material*
- *Perceived by students as a fair representation of achievement*
- *Reflects the skills & topics emphasized in the course*
- *Reveals misconceptions*
- *Feedback to students*
- *Prepare students for subsequent course, let the faculty make a judgment about whether they're ready*
- *Encourage/motivate students to study*
- *Covers aspects from the whole course, emphasize importance*

How can we tell if an exam has met its goals?

Two useful metrics to assess your questions:

- **Difficulty:** Look at the average score for each question – any surprises?
- **Discrimination:** How well does the score on an individual question (item) correspond to performance on the exam?
  - Simple: Calculate question score for each quartile
  - More complicated: Item analysis statistics



Above: Three questions with differing discrimination patterns.

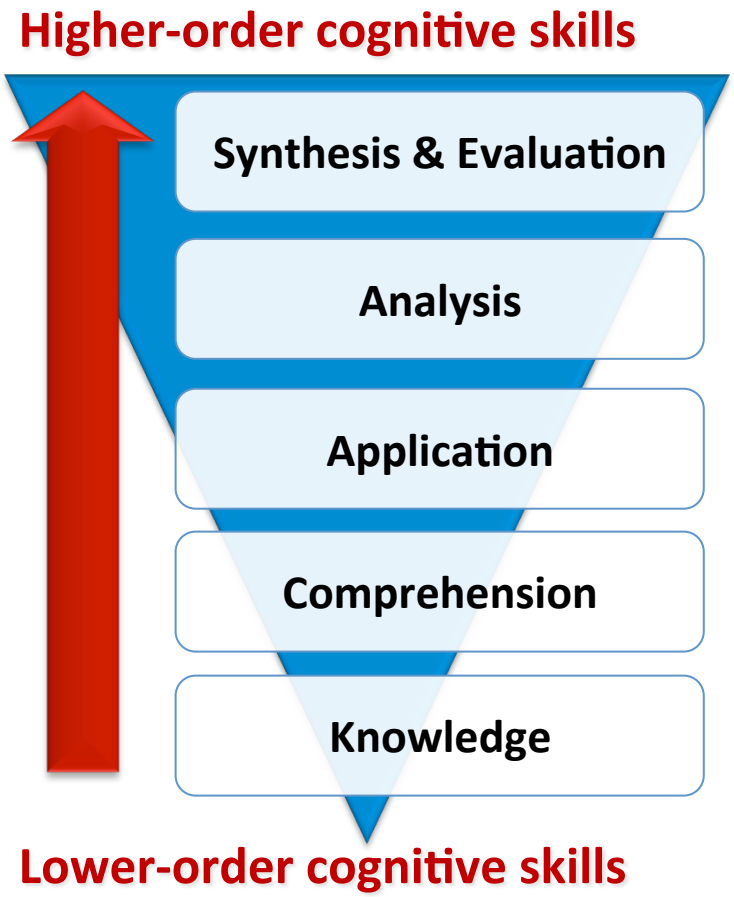
Susan Eaves & Bradley Erford. The Purpose of Item Analysis
http://www.education.com/reference/article/item-analysis/

How can we tell if an exam has met its goals?

For questions with low achievement, or undesirable discrimination, consider difficulty of:

- **Interpretation:** Is interpretation a goal for the question? If not, simplify.
- **Transfer:** new contexts, new ways of using concepts?
- **Level of cognitive (thinking) skills:**
  - Compare exam questions with Learning Goals and practice materials
  - Does a question require integration of 2 or more concepts?

Bloom's Taxonomy, 1956



Determining the cognitive level of exam questions: Tools available

Q1. Are students reproducing something (explanations, definitions, graphs, etc.) that they have seen or heard in course material?

Yes – Go to Q2.
No – Go to Q4.

Q2. To answer the question, are students repeating nearly exactly what they have heard or seen in class materials (including lecture, textbook, lab, homework, clicker, etc.)?

Yes → SEE RECALL

No – Go to Q3.

Q3. Are students demonstrating a conceptual understanding by putting the answer in their own words, matching examples to concepts, representing a concept in a new form (words to graph, etc.), etc.?

Yes → SEE COMPREHENSION

No – GO BACK to Q1. If you are sure the answer to Q1 is yes, the question should fit into RECALL or COMPREHENSION.

J. Casagrand & K. Semsar (UC Boulder). Don't have a pre-post test, use a Bloom's rubric! The development and validation of a rubric for "blooming" assessments to measure student learning. http://www.colorado.edu/sei/documents/publications/IPHY/Blooms-rubric.pdf

What can we learn from an exam once it's over?

Concepts & Skills requiring additional focus or practice:

- Question-by-question performance
- Student responses to open-ended questions
  - Errors of question interpretation
  - Errors in reasoning
  - Misconceptions

What can we learn from an exam once it's over?

Notes for designing future exams:

- **Test equating:** reuse subset of questions to compare student performance year-to-year
- **Item design:**
  - Improve questions with that don't meet goals for difficulty or discrimination
  - Structure future questions to capture different stages of progression towards expert-like thinking
- **Ease of Marking:** can the question be reformatted to reduce ambiguity or save time?

Resources

CWSEI website: http://www.cwsei.ubc.ca/resources/index.html

- "Assessments that support student learning" 2-pager
- "Course alignment" 2-pager

How to prepare better multiple-choice test items (Handbook)

http://testing.byu.edu/info/handbooks/betteritems.pdf

"Blooming" tool for evaluating cognitive level of exam questions: Janet Casagrand and Katharine Semsar (UC Boulder)

- http://www.colorado.edu/sei/documents/publications/IPHY/Blooms-rubric.pdf