



Using Rasch Modeling to Analyze Standards-Based Assessment Items

Cari Herrmann Abell (Research Associate), George DeBoer (PI), & Jo Ellen Roseman (Co-PI), AAAS Project 2061

Background:

Project 2061 is designing a process for creating clusters of multiple choice assessment items to test middle school students' understanding of science ideas from AAAS's Benchmarks for Science Literacy and the National Research Council's National Science Education Standards. Items are being developed in 16 topic areas from the life, physical, and earth sciences; mathematics; and the nature of science. Distractor-driven multiple-choice items make use of common misconceptions as distractors, which allows users to simultaneously test for correct science ideas and documented misconceptions. Items expect students to make use of a variety of cognitive skills including: recognizing the truth of scientific facts and principles, and using the targeted ideas to explain, predict, and analyze phenomena. In this poster, we present the steps in our item development process and a summary of how we use Rasch Modeling to analyze the items.

Steps in Item Development:

(This poster focuses on the steps highlighted below.)

- Select target content standards.
- **Tease apart content standards into finer-grained key ideas.**
- Create assessment maps that visually portray the relationships among target ideas.
- Write clarification statements for each key idea that define the boundaries of knowledge that will be tested.
- Review the literature on student learning to identify misconceptions that students may have and that can be used as distractors in test questions.
- Interview students to fill in gaps in the research literature regarding ideas students have.
- Develop assessment tasks using the criteria of necessity and sufficiency to align items to content standards and improve construct validity of items by paying attention to comprehensibility, test-wisness, and appropriateness of task context.
- Review and revise items internally.
- Pilot test items with a sample of approximately 150 students per item.
- Formally review items using teams of external and internal reviewers.
- **Use Rasch modeling to compare the difficulty of items and determine a discrimination measure for each answer choice.**
- Revise items based on reviewer comments and results of pilot testing.
- **Field test with a national sample of 2000 students per item.**
- Analyze field test results and upload items and item characteristics to our Web site.

Identify Key Ideas Based on Content Standards

The ideas presented here are based on Chapter 4, Section D, of Benchmarks for Science Literacy (BSL) (AAAS, 1993) and Physical Science Content Standard B of the National Science Education Standards (NSES) (NRC, 1996).

- AM-A**: All matter is made up of atoms.
- AM-B**: All atoms are extremely small.
- AM-E**: All atoms and molecules are in constant motion.
- AM-I**: For any single state of matter, the average speed of the atoms or molecules increases as the temperature of a substance increases and decreases as the temperature of a substance decreases.
- AM-G**: For any single state of matter, changes in temperature typically changes the average distance between atoms or molecules. Most substances or mixtures of substances expand when heated and contract when cooled.
- AM-F**: There are differences in the spacing, motion, and interaction of atoms and molecules that make up solids, liquids, and gases.
- AM-H**: Changes of state can be explained in terms of changes in the arrangement, motion, and interaction of atoms and molecules.
- SC-A**: A pure substance has characteristic properties, such as density, a boiling point, and solubility, all of which are independent of the amount of the substance and can be used to identify it.
- SC-C**: Many substances react chemically in predictable ways with other substances to form new substances with different characteristic properties.
- SC-D**: When substances interact to form new substances, the atoms that make up the molecules of the original substances rearrange into new molecules.
- SC-G**: Whenever substances interact with one another, regardless of how they combine or break apart, the total mass remains the same.
- SC-H**: Whenever atoms interact with each other, regardless of how they are arranged or rearranged, the number of each kind of atom stays the same and, therefore, the total mass stays the same.

Field Test Items

Revised items are field tested on a national sample of approximately 2000 students to:

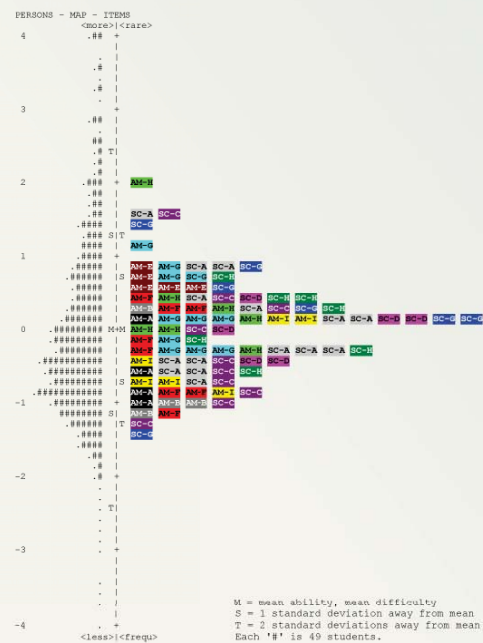
- determine psychometric properties of the items and clusters of items;
- construct difficulty scores, item discrimination indices, and differential item functioning scores; and
- describe student understanding of the targeted science ideas.

Description of the Sample Tested

The results presented here are from two field tests involving over 7000 middle school students in grades 6 through 8 that were administered in the Spring of 2007 and the Spring of 2008. Each item was taken by approximately 2000 students. Additional testing was done during the Fall of 2009 at a highly selective private boarding school located in the Northeast with over 500 students in grades 9 through 12. Students at two universities (a public university in a southern state and a public university in a northeastern state) were also tested in 2008. These students included 474 students who had taken high school chemistry and were enrolled in a college level introductory chemistry course but who had not yet had any instruction at the college level, and over 1200 students who had had at least one semester of college level chemistry instruction.

Rasch Modeling

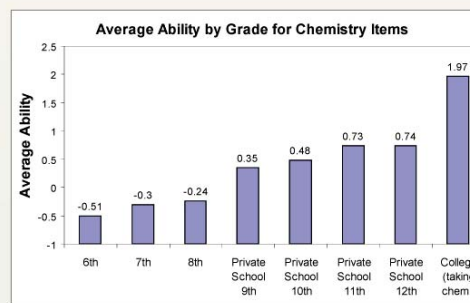
Rasch modeling is used to estimate and compare item difficulty and the popularity of each answer choice for students of differing ability. The item - person map below shows the distribution of student abilities on the left and the item difficulties on the right. When item difficulty and person ability match, the person has a 50% chance of answering the item correctly. Item difficulties are shown for all 91 chemistry items in our field tests. The items are color coded by idea to illustrate the range of difficulties for items in each idea cluster.



The table below lists the key ideas in order of difficulty based on the average of the item measures.

Idea	Average Measure
AM-B	-0.69
AM-A	-0.58
AM-I	-0.40
AM-F	-0.28
SC-C	-0.25
SC-D	-0.03
SC-A	0.04
SC-H	0.12
AM-G	0.25
SC-G	0.34
AM-H	0.35
AM-E	0.64

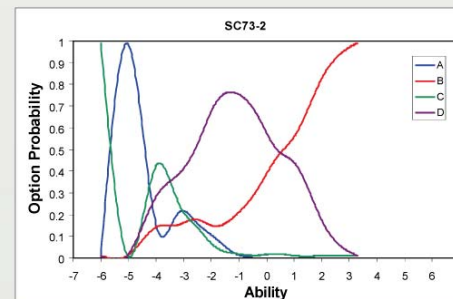
The bar graph shows the average ability of the students in each of the different grade levels as measured by the set of items. The larger increases between middle school and ninth grade and between high school and college can be attributed to the greater selectivity of the sample at those points.



Distractor Analysis

Option probability curves show the probability of selecting each answer choice as a function of student ability. For traditional multiple choice items, the curve for the correct answer typically increases monotonically with increasing ability and the curve for the set of distractors typically decreases monotonically with increasing ability (Haladyna, 1994). In dichotomous scoring, the incorrect answer choices are lumped together and no attempt is made to determine the difference in discriminating ability of those incorrect answers. However, it has also been shown that, for distractor-driven items like the ones we developed, the curves do not match the monotonic behavior of traditional items (Sadler, 1998). The unique shape of the curves for distractor-driven items allows for a more precise analysis. Examining the curves enables us to see if the shape of the curves matches our expectations or if there is something unusual that could indicate a structural problem with an item. The shape of the curves also may be suggestive of a hierarchy of misconceptions which disappear in sequence as students become more knowledgeable about a topic either through out-of-school experience or through formal instruction.

Example 1: The Nature of Chemical Reactions



Item SC73-2

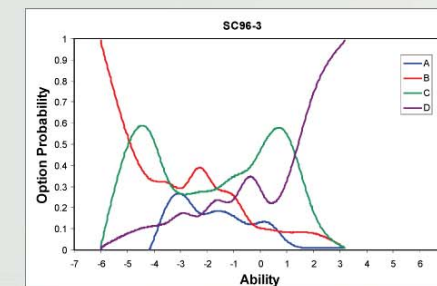
Which of the following statements about chemical reactions is TRUE?

- A. All chemical reactions are dangerous. (8%)
- B. Some chemical reactions can be reversed. (24%)*
- C. Chemical reactions occur only in a laboratory. (7%)
- D. Chemical reactions require starting with at least two substances. (61%)

Copyright © 2008 AAAS Project 2061

This pattern of response by ability is consistent with our expectations. It is reasonable to assume that prior to instruction students' ideas about chemical reactions are based largely on impressions they gain from the popular media. These include the idea that chemical reactions are dangerous and the idea that chemical reactions occur only in a laboratory. As students gain more familiarity with actual chemical reactions during instruction, many of the reactions they encounter involve two starting materials. As instruction progresses into the upper grades, there are increasingly more opportunities for students to learn that at least some chemical reactions are reversible. It is unlikely that students would draw that conclusion from out-of-school experience alone.

Example 2: Characteristic properties of substances



Item SC96-3

You have two balloons. Balloon 1 is filled with a gas, and Balloon 2 is filled with a gas. The gases are pure substances. What could you do to help decide if the gases in the balloons are the same substance?

- A. Compare the temperature of the gas in Balloon 1 to the temperature of the gas in Balloon 2. (17%)
- B. Compare the volume of the gas in Balloon 1 to the volume of the gas in Balloon 2. (27%)
- C. Compare the weight of the gas in Balloon 1 to the weight of the gas in Balloon 2. (32%)
- D. Compare the odor of the gas in Balloon 1 to the odor of the gas in Balloon 2. (23%)*

Copyright © 2008 AAAS Project 2061

The curve for the answer choice in which students say that you can tell whether the gases in the two balloons are the same substance by comparing the weights of the gases (C) is unusual because it shows two distinct peaks, one at very low ability levels and the other at relatively high ability levels. We speculated that some of the more able students may have assumed that the balloons were the same size and, therefore, had the same volume. If this were the case, then comparing the weights would be a very effective, and probably the best, way to decide if the gasses in the balloons were the same substance because equal volumes of different substances usually have different weights. This is a case where the shape of the curve quickly revealed a structural problem with the item that could be corrected by saying that the balloons are different sizes.

Conclusions

Rasch Modeling was used to analyze the results from a set of 91 multiple choice assessment items aligned to middle school ideas about chemistry. For this set of items, we found that students had the most difficulty with ideas about atomic motion, changes of state, conservation of mass, and thermal expansion. The students were most successful with items testing the ideas that all matter is made up of atoms and atoms are extremely small. We also found a steady increase in understanding of chemistry from 6th grade students to college students. We examined the option probability curves for each item, which revealed hierarchies of misconceptions. Additionally, the unusual shape of one of the distractor probability curves for an item indicated a structural problem with that item.

References

AAAS (1993). Benchmarks for science literacy. New York: Oxford University Press.

Haladyna, T. M. (1994). Developing and Validating Multiple-Choice Test Items. Hillsdale, NJ: Erlbaum.

Linacre, J.M. (2006). WINSTEPS Rasch measurement computer program. Chicago: Winsteps.com.

NRC (1996). National science education standards. Washington, DC: National Academy Press.

Sadler, P. M. (1998). Psychometric Models of Student Conceptions in Science: Reconciling Qualitative Studies and Distractor-Driven Assessment Instruments. Journal of Research in Science Teaching, 35, 265-296.

Acknowledgments

Research Team
Abigail Burrows ♦ Jean Flanagan ♦ David Pollock ♦ Brian Sweeney ♦ Jill Wertheim ♦ Ted Willard

NSF Grant No. ESI-0352473