Aligning Assessment to Content Standards:

Applying the Project 2061 Analysis Procedure to Assessment Items in School Mathematics

George E. DeBoer, AAAS Project 2061; Paul Ache, Kutztown University of Pennsylvania

The Study:

This study was designed to determine the effectiveness of a procedure to improve the alignment of mathematics assessment items to targeted state content standards.

The study was conducted on over 100 released items from a single state in the Northeast.

The items were analyzed using a procedure developed by Project 2061 of AAAS. There were three broad criteria on which the items were analyzed:

Content Alignment: Is the knowledge specified in the content standard needed to answer correctly or can the correct answer be obtained in some other way? Is the knowledge specified in the content standard enough by itself to make a satisfactory response or is additional knowledge or skill needed as well?

Item Efficiency: Is there anything in the item, not related to understanding the targeted content standard, that might interfere with a student's ability to respond correctly? Issues include comprehensibility, appropriateness of the task context, and "guessability."

Plausibility of Answer Choices: Are all answer choices plausible and related to the ideas being tested? For example, are distractors related to students' misconceptions and commonly held beliefs?

Teams of analysts produced written profiles that described each item's alignment with the targeted content standard and provided suggestions for revision. Items were revised on the basis of the analysis criteria, not on an examination of student responses on the original items.

Revised and original items were given to students who were asked to show their work, explain how they obtained their answer, and to indicate if anything about the item was confusing. Two forms of a test were created for each grade. Half of the items on each form were original and half were revised. Test forms were distributed randomly in each class. Data were analyzed to determine the impact that revisions had on improving the match between students' answer choices and their written explanations. The study provided information about the validity of this analysis procedure for improving the alignment of assessment items to content standards. Data presented here are for six items that were field tested with 259 eleventh grade students.

Results:

Comparing Original and Revised Items

	Provided Explanations (%)		False Neg/Pos (%)		Confused by Wording (%)		Confused Total (%)		Difficulty (% correct)	
Item	*Orig.*	*Rev.*	*Orig.*	*Rev.*	*Orig.*	*Rev.*	*Orig.*	*Rev.*	*Orig.*	*Rev.*
1	81.5	77.1	18.6	1.9	9.3	0.9	9.3	12.0	88.2	85.0
2	70.0	73.9	17.3	21.6	14.3	8.0	22.4	19.3	52.1	60.5
3	74.8	80.6	13.5	0.0	0.0	0.0	39.3	17.0	26.9	41.0
4	71.2	64.7	3.0	9.1	13.1	29.9	25.2	59.7	57.6	28.6
5	72.7	85.0	10.2	8.8	2.3	2.7	22.7	27.4	48.8	48.8
6	73.4	73.8	7.8	2.2	3.9	5.5	12.7	17.8	66.9	58.2
Mean	**73.9**	**75.9**	**11.7**	**7.3**	**7.2**	**7.8**	**21.9**	**25.5**	**56.8**	**53.7**

Conclusions:

- Whenever possible, the procedure should make use of student response data before items are revised. The purpose of the analysis procedure is to reduce the number of student responses that do not accurately reflect what they know and can do. Some of the factors that lead students to answer correctly when they do not have the required knowledge and incorrectly when they do are not apparent until student responses are examined.
- When students are asked to provide explanations for their answers or to show their work, approximately 75% of them do so. The comments that they make are helpful for determining if the answer they selected on a multiple choice test is consistent with their understanding as shown in their work and explanations.
- When students are asked if anything in an assessment task is confusing to them, they answer in three ways: (1) they identify specific mathematics content that they do not understand; (2) they identify specific wording or aspects of the structure of the item that is confusing; and (3) they offer comments about being confused in general, without specifying what was confusing to them. Most answers to this question are about content confusion, although in a small but significant number of cases the students provide specific information about wording that is helpful when revising items.

The research reported here was supported by the National Science Foundation (NSF Grant #9819018). Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect those of the NSF.

For more information: American Association for the Advancement of Science (AAAS) Project 2061
www.project2061.org

Item 1.

Targeted Content Standard:
Use operations (e.g. opposite, reciprocal, absolute value, raising to a power, finding roots, finding logarithms.)

Original: Which of the following represents the largest value?

A. 10^3
B. $(5 + 5) \times 10$
C. $\frac{10^8}{10^5}$ *(correct response)*
D. $10^3 \times (10)^2$

Revised: Which of the following expressions represents the value 10,000?

A. $10^4 - 10^0$
B. $10^2 + 10^2$
C. $\frac{10^8}{10^7}$
D. $10^4 \times 10^0$ *(correct response)*

What we found:

On the original item, eight of 97 students were able to evaluate the exponents in all answer choices correctly except for the expression $10^3 \times (10)^2$, which they interpreted to be $(10^3 \times 10)^2$, and said that the form of the expression confused them. The expression $10^3 \times (10)^2$ is written in a non-standard form. This was changed in the revised item and ***the number of students indicating wording confusion dropped accordingly.***

Item 2.

Targeted Content Standard:
Apply ratio and proportion to mathematical problem situations involving distance, rate, and similar triangles.

Original: Kim needs a certain shade of pink paint for a handmade toy. This shade is made by mixing white and red paint in a ratio of 1 to 3. How many fluid ounces of **red paint** would be needed to make 12 fluid ounces of this pink paint?

A. 4 FLUID OUNCES
B. 6 FLUID OUNCES
C. 8 FLUID OUNCES
D. 9 FLUID OUNCES *(correct response)*

Revised: Kim is painting a handmade toy and she needs to mix paint so she can create a certain shade of pink. This shade is made by mixing white paint and red paint in a ratio of 2 ounces to 5 ounces. How many ounces of **red paint** are needed to mix with 100 ounces of white paint to create the right shade of pink?

A. 500 OUNCES
B. 250 OUNCES
C. 40 OUNCES *(correct response)*
D. 20 OUNCES

What we found:

On the original item, 14 of 98 students correctly used equivalent fractions but reversed the order of white and red paint. (The item says "white and red paint in the ratio of 1 to 3.") The order of the ratio was not clarified on the revised item. On the revised item, 16 of 88 students showed that they knew how to calculate equivalent fractions but also reversed the order of the ratio and got the question wrong. ***If student responses had been available before revision, the order of paint in the ratio could have been clarified in the revision.***

Item 3.

Targeted Content Standard:
Develop and use computation concepts, operations and procedures with real numbers in problem-solving situations.

Original: At the start of the month, the counter on the copy machine read 6,583. At the end of the month, it read 82,110. The copies cost $1\frac{1}{2}$ cents a piece. What was the approximate total cost of the copies for this month?

A. $10,000.00
B. $2,200.00
C. $1000.00 *(correct response)*
D. $200.00

Revised: At the start of the month, the counter on the copy machine read 733,296. At the end of the month, it read 758,312. The copies cost $1\frac{1}{4}$ cents a piece. What was the cost of the copies for this month?

A. $312.70 *(correct response)*
B. $333.55
C. $6255.00
D. $31,270.00

What we found:

On the original item, which asked for an approximate cost, nine students got the correct answer even though they made serious miscalculations of the cost of the copies. The answer choices were about an order of magnitude apart so that using just the final reading on the copy machine or misrepresenting the fraction still produced a correct answer. These were considered to be false positives. The revised item asked for an actual cost of the copies, not an approximate value. ***There were no false positives on the revised item.***

Item 4.

Targeted Content Standard:
Develop and use computation concepts, operations and procedures with real numbers in problem-solving situations.

Original: As part of her pay, a real estate agent is given a commission which is a percentage of the sale price. What is her commission on a farm which sold for $1,750,000? The rates are as follows:

Sale Price	Commission
First $500,000	8%
Any portion over $500,000	5%

A. $ 87,500
B. $100,000
C. $102,500 *(correct response)*
D. $140,000

Revised: A real estate agent is paid 8% of sales that are $500,000 or less. She is paid an additional 5% of portions of sales over $500,000. What will her pay be for the sale of a farm for $1,750,000?

A. $ 87,500
B. $102,500 *(correct response)*
C. $125,000
D. $140,000

What we found:

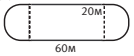
When examining the original item, analysts determined that the ability to interpret a data table went beyond what this item was testing (did not meet the sufficiency criterion) and, therefore, removed it during revision. The result was that ***the item became much more difficult for students*** (57.6% correct on the original item; 28.6% correct on the revised item). In addition, ***students indicated greater confusion on the revised item*** (25.2% confused on the original item and 59.7% confused on the revised item). The data table apparently helped students to see how the commission was divided into two portions, a concept that was difficult for them to grasp otherwise.

Item 5.

Targeted Content Standard:
Estimate, use, and describe measures of distance, rate, perimeter, area, volume, weight, mass and angles.

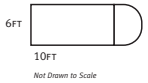
Original: The inside rail of a running track consists of a rectangle with a semicircle at each end as shown in the figure below. Find the approximate area surrounded by the track.

A. 1200 M^2
B. 2456 M^2
C. 160 M^2
D. 1514 M^2 *(correct response)*



Revised: Below is a garden in the shape of a rectangle with a half-circle attached to one end. Find the approximate area of the garden.

A) 35 FT^2
B) 60 FT^2
C) 74 FT^2
D) 88 FT^2



Not Drawn to Scale

What we found:

On the original item, six students were able to answer correctly by either ignoring or not knowing how to calculate the area of the two semicircles. This was possible because the correct answer can be obtained by calculating the area of the rectangle and then adding a little bit more. The same was true for the revised item, for which six students got the correct answer by ignoring the area of the semicircle or by miscalculating and having their answer still be closest to the correct answer. ***If student responses had been available before revision, this could have been taken into account in the revision.***

Item 6.

Targeted Content Standard:
Draw deductive and inductive conclusions within mathematical contexts.

Original: In a group of 10 people, 7 people speak English, 4 people speak French and 2 speak neither of the two languages. How many people speak both languages?

A. 1
B. 2
C. 3 *(correct response)*
D. 4

Revised: In a group of 20 people, 15 people speak English, 7 people speak French and 2 speak neither of the two languages. How many people speak both languages?

A. 2
B. 3
C. 4 *(correct response)*
D. 8

What we found:

On the original item, six students got the correct answer by subtracting the number of people who spoke French from the number who spoke English. These were considered to be false positives. On the revised item it was not possible to get the correct answer in this way and ***the number of false positives was reduced*** accordingly.