The Anatomy of High-Quality Multiple Choice Assessment Items

Research in Mathematics Education

Cassandra Hatfield

Deni Basaraba, Ph.D.
Is this the test to test us for the test to see if we are ready for the test?
Focus on Assessment

<table>
<thead>
<tr>
<th>Discussion Points</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Knowledge Representations</td>
<td>• Understand the anatomy of a test item</td>
</tr>
<tr>
<td>• Levels of Mathematics Proficiency</td>
<td></td>
</tr>
<tr>
<td>• Appropriate Question Stems</td>
<td></td>
</tr>
<tr>
<td>• Including Student Misconceptions in answer choices</td>
<td></td>
</tr>
<tr>
<td>• Importance of technically adequate assessments</td>
<td>• Reliable data decision making</td>
</tr>
<tr>
<td>• Examining student performance: Moving beyond the overall score</td>
<td></td>
</tr>
</tbody>
</table>
## Item Writing Template

<table>
<thead>
<tr>
<th>Grade Level:</th>
<th>TEKS:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge Representation (Circle One):</td>
<td>Foundational</td>
</tr>
<tr>
<td>Cognitive Engagement (Circle One):</td>
<td>Procedural</td>
</tr>
<tr>
<td>Relative Difficulty (Circle One):</td>
<td>Easy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stem (Text and Graphic)</th>
<th>Response Choices</th>
<th>Content/Graphic</th>
<th>Student Misconception Or Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distractor 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distractor 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distractor 3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Approval:**

**Reviewer Comments:**
Considerations for Test Items

• **Knowledge Representations**
  – Target knowledge and skills
  – Bridging knowledge and skills
  – Foundational knowledge and skills

• **Strands of Mathematical Proficiency**
  – Conceptual understanding
  – Procedural fluency
  – Strategic competence
  – Adaptive reasoning
Purpose of Different Types of Assessments

- **Universal Screening**
  - Who is struggling?
  - What level of intensity of support does the student need?
  - 3-Tier Response to Intervention Model

- **Diagnostic Assessments**
  - Why are they struggling?
  - How deep are the gaps in the students understanding?
  - Learning Progressions

- **Skills Inventory**
  - Why are they struggling?
  - What grade level concepts do I need to intervene with to get them back on track?
  - Individual Standards
SMU
Knowledge Representations

Foundational Knowledge and Skills

Bridging Knowledge and Skills

Target Knowledge and Skills
Knowledge Representations: Target Knowledge and Skills

- Grade level knowledge and skills
- Supports **future success in mathematics**
- Often are abstract representations of formal mathematical knowledge, but not always
Knowledge Representations: Foundational Knowledge and Skills

- Knowledge and skills that support the target content and are accumulated from previous learning
- **Prerequisite knowledge and skills** needed to be successful at the target knowledge and skills
Knowledge Representations: Bridging Knowledge and Skills

- Knowledge and skills needed to **connect or support students’ learning** from the foundational knowledge to the target knowledge and skills
  - Often represents an integration of knowledge and skills (may be conceptual or model-based)
- The knowledge and skills that students learn from the teacher or instructional materials
Gracie has 12 books. She has an empty bookshelf with 3 shelves. She puts the same number of books on each shelf, how many books will be on each shelf?

A. 3 books
B. 4 books
C. 6 books
D. 12 books

Correct answer: B
Which model could be used to represent $36 \div 9 = \square$?

A. 

B. 

C. 

D. 

Correct answer: C
Curriculum Focal Points are:

- Framework of critical areas of mathematics instruction
- A mathematical theme, not a single TEKS statement

Before writing assessments, it is important to:

- Analyze the focal point description to identify key target skills
- Analyze related TEKS statements
- Synthesize the overlaying skills
Process for Articulating the Content of an Item through knowledge representations

- Become familiar with the TEKS standard (content standard) for which you are writing a test item
- Articulate the **TARGET SKILLS**
- Articulate the **FOUNDATIONAL SKILLS**
- Articulate the **BRIDGING SKILLS**
Knowledge Representations

**Foundational:**
Skills that support the target goal and are accumulated from previous learning

**Target:**
Grade level mathematics reasoning and knowledge

**Bridging:**
Mathematical knowledge needed to connect foundational with abstract mathematical reasoning
Strands of Mathematical Proficiency

- Intertwined Strands of Proficiency

National Research Council (2001)
Conceptual Understanding

- Demonstrate an integrated and functional grasp of mathematical ideas
- Understand specific task as it relates to a whole concept
- Find relationships between pieces of information
- Make connections to similar representations
- Use models and multiple representations (e.g. pictures, numbers, real-life situations, words)
TEKS 6.4E
The student applies mathematical process standards to develop an understanding of proportional relationships in problem situations. The student is expected to:

(E) Represent ratios and percents with concrete models, fractions, and decimals.

Which model has an equivalent ratio of shaded parts to total parts as this rectangle?

A  
B  
C  
D

Correct answer: C
TEKS 3.3D

The student applies mathematical process standards to represent and explain fractional units. The student is expected to:

(D) **Compose and decompose a fraction** $a/b$ with a numerator greater than zero and less than or equal to $b$ as a sum of parts $1/b$.

Which number line shows $\frac{3}{5}$?

A. 

B. 

C. 

D. 

Correct answer: B
Procedural Fluency

- Use formal language or symbolic representations
- Carry out accurate computations
- Follow multiple steps sequentially
- Make proper use of algorithm and properties
TEKS 7.11A

The student applies mathematical process standards to solve one-variable equations and inequalities. The student is expected to:

(A) Model and solve one-variable, two-step equations and inequalities.

Solve for w:

\[
\frac{w + 12}{3} = 20
\]

A. 72  
B. 58  
C. 48  
D. 16

Correct answer: C
TEKS 4.3E
The student applies mathematical process standards to represent and generate fractions to solve problems. The student is expected to:

(E) **Represent and solve addition and subtraction of fractions with equal denominators using objects and pictorial models that build to the number line and properties of operations.**

Mario cuts a pie into 6 equal slices. He eats 1 slice for lunch and 1 slice for dinner. Which equation shows the fraction of a whole pizza that Mario eats?

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$\frac{1}{6} + \frac{1}{6} = \frac{1}{12}$</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>$\frac{1}{6} + \frac{1}{6} = \frac{2}{12}$</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>$\frac{1}{6} + \frac{1}{6} = \frac{2}{6}$</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>$\frac{1}{3} + \frac{1}{3} = \frac{2}{6}$</td>
<td></td>
</tr>
</tbody>
</table>

Correct answer: **C**
Strategic Competence

- Ability to formulate a problem in mathematical terms
- Represent problem solving strategically (verbally, symbolically, graphically, or numerically)
- Identify and use strategy necessary to solve problems effectively (e.g. use the distributive property to solve)
The perimeter of the figure below is 43 units.

Which equation can be used to solve for the variable, $g$?

A. $7.5g + 13 = 43$
B. $7g + 13 = 43$
C. $5.5g + 13 = 43$
D. $5g + 13 = 43$

Correct answer: A
Samantha takes $140 to the toy store. She buys a doll house and has $58 left over. Which number sentence can be used to find the cost of the doll house?

A \[ \square - 140 = 58 \]

B \[ 140 + 58 = \square \]

C \[ 140 + \square = 58 \]

D \[ 140 - \square = 58 \]

Correct answer: D
Adaptive Reasoning

- Think logically about a problem, which requires reflecting on various approaches to solve a problem and deductively selecting an approach
- Rationalize and justify strategies
- Appropriately explain a procedure or concept
Adaptive Reasoning

TEKS 3.3H

The student applies mathematical process standards to represent and explain fractional units. The student is expected to:

(H) Compare two fractions having the same numerator or denominator in problems by reasoning about their sizes and justifying the conclusion using symbols, words, objects, and pictorial models.

Why is $\frac{2}{3} > \frac{2}{7}$?

A. 2 is equal to 2 and thirds are larger than sevenths.
B. 2 is equal to 2 and thirds are smaller than sevenths.
C. 2 is equal to 2 and 3 is smaller than 7.
D. 2 is equal to 2 and 3 is greater than 7.

Correct answer: A
TEKS 6.2B
The student applies mathematical process standards to represent and use rational numbers in a variety of forms. The student is expected to:

(B) Identify a number, its opposite, and its absolute value.

Adam says that the absolute value of -7 is 7. Why is he correct?

A. 7 is the distance from -7 to 0 on the number line.

B. 7 is a prime number with only two factors.

C. 7 is an odd number, so it is not divisible by 2.

D. 7 is the square root of 49, which is a perfect square.

Correct answer: A
Level of Difficulty

- **Easy**
  - Basic Knowledge
  - Skills that are familiar to students
  - Sometimes conceptually based

- **Medium**

- **Difficult**
  - Skills that are peripheral to curriculum
  - Not all students will have mastered these

(Cheng, 2005)
In your packet of materials is a 4 x 3 matrix with the 4 strands of mathematical proficiency along the top and 3 levels of difficulty along the left side.

Take a blue set of cards. They should have a number in the corner, ranging from 1 to 12.

As you read each item, determine the strand of mathematical proficiency and relative level of difficulty.

Write the number associated with the item in the appropriate cell.
### Item Match – Answer Key

<table>
<thead>
<tr>
<th></th>
<th>Procedural</th>
<th>Conceptual</th>
<th>Strategic</th>
<th>Adaptive</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Easy</strong></td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td><strong>Medium</strong></td>
<td>10</td>
<td>6</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td><strong>Difficult</strong></td>
<td>8</td>
<td>11</td>
<td>9</td>
<td>2</td>
</tr>
</tbody>
</table>
Consider your purpose for giving the assessment when using these considerations

• **Knowledge Representations**
  – Target knowledge and skills
  – Bridging knowledge and skills
  – Foundational knowledge and skills

• **Strands of Mathematical Proficiency**
  – Conceptual understanding
  – Procedural fluency
  – Strategic competence
  – Adaptive reasoning
Assessment Item Development
• Avoid the complex multiple-choice format. (i.e., A and D, B and C).
• Use plain language. Avoid conditional phrases (if…, then…).
• Keep the language of the stem and response options at the appropriate grade level.
• Minimize examinee reading time in phrasing each item.

(Haladyna, 2004)
Stem Development

- State the stem in question form. Minimize use of completion form. When using the completion format, do not leave a blank for completion in the beginning or middle of the stem.

- Include only the material needed to make the problem clear. Avoid extraneous information.

- Word the stem positively; avoid negative phrasing. If an item must be stated negatively, underline or capitalize the negative word.

- Keep all essential information in the stem. Items that require students to read and evaluate each response option prior to selecting an answer increase the cognitive load required.

(Haladyna, 2004)
General Item-Writing (Content)

- Base each item on important content to learn; avoid trivial content.
- Keep the content of each item independent from content of other items on the test.
- Avoid cuing one item with another; keep items independent of one another.
- Avoid items based on opinions.
- Develop items that measure higher-level thinking.
- Avoid potentially insensitive content or language.

(Haladyna, 2004)
Response Development

• Make all distractors plausible.
  – If you’re interested in obtaining more information about students’ understanding, create distractors that represent common misconceptions may have about the content being assessed

• Keep all options in an item homogenous in content and grammatical structure.

• Keep the length of options brief and fairly consistent.

• Phrase options positively, not negatively.

(Haladyna, 2004)
Dan enjoys going to the park. The park has lots of trees and a pond. Dan goes to the park at 10:15 a.m., plays on the swings, slides down the slide, and runs around the pond. He is hungry when he finally makes it home 1 hour 55 minutes later. What time did he arrive at home?

A. 12:00 p.m.
B. 12:10 p.m.
C. 8:20 a.m.
D. 6:00 p.m.
Sue has a box of 2 stars and 2 circles. She wants to make her box proportional to Michelle’s box of 6 stars and 3 circles. How many stars does she need to add to her box to make it proportional to Michelle’s box?

A. 6  
B. 0  
C. 2  
D. 20
Data-Driven Decision Making
Purpose of Different Types of Assessments

**Universal Screening**
- Who is struggling?
- What level of intensity of support does the student need?
- 3-Tier Response to Intervention Model

**Diagnostic Assessments**
- Why are they struggling?
- How deep are the gaps in the students understanding?
- Learning Progressions

**Skills Inventory**
- Why are they struggling?
- What grade level concepts do I need to intervene with to get them back on track?
- Individual Standards
Decision Making Checklist for Assessment Development

Teachers should agree to analyze the assessments around the same set of criteria. The decisions should be directed toward:

✓ Validation about the appropriateness of the assessment
✓ Ensuring the assessment is congruent to the stated mastery objective and/or state or district standards
✓ Consistency of opinion about the assessment and evaluation of the work
✓ Adjustments in teacher directions and support for all students

(Rutherford, 2008)
Barriers to Decision Making

- Misrepresentation Of Construct
- Misrepresentation Of Test/Item Format
- Under-Representation Of the Construct

Targeted Construct For Making Interpretations
Using Data to Inform Instruction: Overall Student Performance

- One of the most common ways to examine student performance data is by making normative comparisons of their overall performance:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Teacher</th>
<th>Student</th>
<th>Number Correct</th>
<th>Percent Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Richardson</td>
<td>Swan, B.</td>
<td>20/40</td>
<td>50%</td>
</tr>
<tr>
<td>5</td>
<td>Richardson</td>
<td>Black, J.</td>
<td>22/40</td>
<td>55%</td>
</tr>
<tr>
<td>5</td>
<td>Richardson</td>
<td>Cullen, E.</td>
<td>21/40</td>
<td>52.5%</td>
</tr>
<tr>
<td>5</td>
<td>Richardson</td>
<td>Newton, M.</td>
<td>20/40</td>
<td>50%</td>
</tr>
<tr>
<td>5</td>
<td>Hatfield</td>
<td>Everdeen, K.</td>
<td>37/40</td>
<td>92.5%</td>
</tr>
<tr>
<td>5</td>
<td>Hatfield</td>
<td>Hawthorne, G.</td>
<td>36/40</td>
<td>90%</td>
</tr>
<tr>
<td>5</td>
<td>Hatfield</td>
<td>Meelark, P.</td>
<td>37/40</td>
<td>92.5%</td>
</tr>
<tr>
<td>5</td>
<td>Hatfield</td>
<td>Abernathy, H.</td>
<td>35/40</td>
<td>87.5%</td>
</tr>
</tbody>
</table>
Using Data to Inform Instruction: Digging a Little Bit Deeper

- Although students’ overall performance may be similar, this does not necessarily mean that they have similar levels of mastery on the assessed content:

<table>
<thead>
<tr>
<th>Student</th>
<th>Number of Items Correct</th>
<th>Total Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number Line Structure</td>
<td></td>
</tr>
<tr>
<td>Swan, B.</td>
<td>7/10</td>
<td>20/40</td>
</tr>
<tr>
<td>Black, J.</td>
<td>5/10</td>
<td>22/40</td>
</tr>
<tr>
<td>Cullen, E.</td>
<td>8/10</td>
<td>21/40</td>
</tr>
<tr>
<td>Hale, J.</td>
<td>5/10</td>
<td>20/40</td>
</tr>
<tr>
<td></td>
<td>Magnitude as Distance</td>
<td></td>
</tr>
<tr>
<td>Swan, B.</td>
<td>6/10</td>
<td></td>
</tr>
<tr>
<td>Black, J.</td>
<td>7/10</td>
<td></td>
</tr>
<tr>
<td>Cullen, E.</td>
<td>5/10</td>
<td></td>
</tr>
<tr>
<td>Hale, J.</td>
<td>7/10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part-to-Whole Relationships</td>
<td></td>
</tr>
<tr>
<td>Swan, B.</td>
<td>5/10</td>
<td></td>
</tr>
<tr>
<td>Black, J.</td>
<td>4/10</td>
<td></td>
</tr>
<tr>
<td>Cullen, E.</td>
<td>3/10</td>
<td></td>
</tr>
<tr>
<td>Hale, J.</td>
<td>3/10</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Unit Fractions</td>
<td></td>
</tr>
<tr>
<td>Swan, B.</td>
<td>2/10</td>
<td></td>
</tr>
<tr>
<td>Black, J.</td>
<td>6/10</td>
<td></td>
</tr>
<tr>
<td>Cullen, E.</td>
<td>5/10</td>
<td></td>
</tr>
<tr>
<td>Hale, J.</td>
<td>5/10</td>
<td></td>
</tr>
</tbody>
</table>
Similarly, even though students may have responded correctly to the same number of items within a given sub-level, again this does not necessarily mean they have mastered the same skills or have the same level of proficiency with the targeted skill.

<table>
<thead>
<tr>
<th>Student</th>
<th>Part-to-Whole Relationships (Items 1-10)</th>
<th>Number Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1  2  3  4  5  6  7  8  9  10</td>
<td></td>
</tr>
<tr>
<td>Swan, B.</td>
<td>1  0  1  1  0  1  0  0  0  1</td>
<td>5/10</td>
</tr>
<tr>
<td>Black, J.</td>
<td>1  1  0  0  1  1  0  0  0  0</td>
<td>4/10</td>
</tr>
<tr>
<td>Cullen, E.</td>
<td>1  1  0  0  0  1  0  0  0  0</td>
<td>3/10</td>
</tr>
<tr>
<td>Hale, J.</td>
<td>1  0  1  0  0  1  0  0  0  0</td>
<td>3/10</td>
</tr>
</tbody>
</table>

0 = Incorrect
1 = Correct
Using Data to Inform Instruction: How Far Down the Rabbit Hole Should I Go?

Clearly, the level at which you can examine student performance can become increasingly fine-grained. For example, once you start looking at item-level performance you can examine any of the following attributes:

- Whether students selected one distractor more frequently than the others (e.g., Did all students who got the item incorrect select the first distractor?)
- Whether students responded correctly to items targeting a specific level of proficiency (e.g., Did students get all of the items targeting procedural fluency correct?)
- Whether students consistently selected the incorrect response for items targeting a specific level of proficiency (e.g., Did students consistently get items targeting strategic competence or adaptive reasoning incorrect?)
Responses to these questions can then be used to further guide instructional planning. For example:

Did students who got the item incorrect select the same distractor?

- **Yes**
  - Target instruction to address the misconception or error in students’ thinking represented by the distractor

- **No**
  - Look for other similarities among the responses selected
Using Data to Inform Instruction: How Far Down the Rabbit Hole Should I Go?

Did students respond correctly to all items targeting a certain level of cognitive complexity?

Yes

Consider varying the examples used during instruction and on assessments to include items that require more complex cognitive processing.

No

Look for other similarities among the responses selected.
Now it’s time to consider the **BIG** question….

**Do I need to look at the data this closely for every student in my classroom?**

Our response….

Probably not. For students who are on track, monitoring their progress using the overall score is probably enough. For students who are struggling, however, digging a bit deeper into the data to try and figure out *why* they are struggling may be very useful.

Understanding *why* students are struggling can be used to help us target our instructional efforts to meet students’ needs.
Some Final Thoughts & Take-Aways

• **Assessment and instruction should be considered together** – performance on assessments can inform instruction and assessments can be specifically designed to provide students with opportunities to demonstrate what they’ve learned during instruction.

• When considering the level of knowledge represented while designing a test item, start with the outcome (**target knowledge**) in mind. Then consider what **foundational knowledge** the student needs and the **bridging knowledge** that will help the student acquire the **target knowledge** or skill.
Some Final Thoughts & Take-Aways

• When designing tests or assessments for use in your classroom, be sure to include items that target multiple levels of cognitive engagement.

• Revisit the guidelines for item development as often as needed to ensure that the items you write provide students with the best opportunity possible to demonstrate their knowledge and understanding of the content.

• Don’t hesitate to look beyond the overall test score to students’ performance on groups of items or individual items to try and understand what instructional supports you can provide to support students’ learning.


SMU Research in Mathematics Education- RME

2016 RME Research-to-Practice Conference
February 19, 2016

Cassandra Hatfield  
chatfield@smu.edu

Deni Basaraba, Ph.D.  
dbasaraba@smu.edu