Quantifying and Improving the Accessibility of Tests and Test Items

Peter A. Beddow

VANDERBILT
Peabody College
1) Definition of accessibility;

2) Test Accessibility and Modification Inventory (TAMI);
   a) Evolution and application of the instrument;
   b) Guiding influences;
   c) Content and Structure of the TAMI and TAMI Accessibility Rating Matrix (ARM);

3) Directions for future research.
• **Accessibility** is the extent to which a product, environment, or system eliminates barriers and permits equal access to all components and services for all individuals. (Beddow, Kettler, & Elliott, 2008)

• **Test accessibility** is the extent to which a test and its constituent item set permits the test-taker to demonstrate knowledge of the target construct. Thus, an accessible test:

1. Eliminates barriers;

2. Permits equal access to all components and features for all test-takers; and

3. Yields scores from which subsequent inferences do not reflect error that is the result of incomplete test-taker access.
• Additionally, accessibility involves an interaction between characteristics of the test and individual test-taker characteristics.

• An item may permit one individual to access the target construct with minimal effort, whereas for another individual, the same item may require the expenditure of essential cognitive resources to gain access the target construct.

• Both individuals may be equally knowledgeable of the tested content, but accessibility issues may preclude one from demonstrating what he or she knows.

• The evaluation of test and test item accessibility requires familiarity not only with the tested material but also with the target population of a test.
Test-Taker Characteristics

• In the case of tests for students with special needs (e.g., students identified with disabilities), characteristics of the test-taker population include:
  • Inattention;
  • Organizational difficulties;
  • Poor reading fluency;
  • History of below-proficient test performances;
  • Low test self-efficacy.
• Just as validity is discussed in terms of degrees, a test is never deemed entirely accessible or entirely inaccessible.

• Most - if not all - tests can be improved with respect to accessibility.

• The TAMI and TAMI Accessibility Rating Matrix (ARM) can be used to guide determinations about which aspects of test items may be improved to enhance their accessibility.

• Ultimately, accessibility evaluations should be based on a firm understanding of universal design principles, cognitive load theory, and research on test and item development.
5. Mr. Jameson is a salesman who lives in Knoxville, Tennessee. He travels the distance from his home to Jackson and back twice a month. He also travels to Chattanooga and back home twice a month. If the distance from Knoxville to Jackson is 295 km and the distance from Knoxville to Chattanooga is 95 km, how many km does he drive on these trips each month?

- A. 670 km
- B. 780 km
- C. 1,560 km
- D. 14,103 km

<table>
<thead>
<tr>
<th>AA-MAS Eligibility</th>
<th>Response Frequency</th>
<th>Item Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not eligible</td>
<td>A 16% B 50% C* 30% D 4%</td>
<td>p 0.30 DOK 2 R 8.7 # Words 66</td>
</tr>
<tr>
<td>Eligible</td>
<td>A 25% B 36% C 21% D 15%</td>
<td>p 0.21</td>
</tr>
</tbody>
</table>

*key; R = Readability Score
<table>
<thead>
<tr>
<th>Project</th>
<th>Instrument</th>
<th>Purpose</th>
<th>Influences</th>
</tr>
</thead>
</table>
| Consortium for Alternate Assessment          | Item Accessibility and Modification Guide (IAMG) | Checklist to sensitize assessment professionals from four states (AZ, ID, IN, HI) to research and theoretical principles for modifying a set of 8th-grade reading and math items. | • NCEO guidance (e.g., Thompson, Johnstone, & Thurlow, 2002; Johnstone, Thurlow, Moore, & Altman, 2006)  

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Influences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modification and Experimental Studies (CAAVES)</td>
<td>• Item-writing research (e.g., Haladyna, Downing, &amp; Rodriguez, 2002)</td>
</tr>
</tbody>
</table>
TAMI: Theory and Development
• The Test Accessibility and Modification Inventory (TAMI; Beddow, Kettler, & Elliott, 2008) was developed as a decision-making tool to facilitate the analysis of new and existing tests and test items with the purpose of enhancing their accessibility.

• The TAMI was influenced by four primary areas of study:

  1) Universal design principles;

  2) Cognitive load theory;

  3) Research on test and item development; and

  4) Guidance on web and computer accessibility.
• Universal design, as defined in the Assistive Technology Act (P.L. 105-394, 1998), is “a concept or philosophy for designing and delivering products and services that are usable by people with the widest possible range of functional capabilities, which include products and services that are directly usable (without requiring assistive technologies) and products and services that are made usable with assistive technologies” (§3(17)).

• The use of universal design principles reduces the need for many specific kinds of assistive technology devices and assistive technology services by building in accommodations for individuals with disabilities before rather than after production. The use of universal design principles also increases the likelihood that products (including services) will be compatible with existing assistive technologies. These principles are increasingly important to enhance access to information technology, telecommunications, transportation, physical structures, and consumer products (PL105-394(§3(10)); emphasis added).
“Considering the wide variety of different variables that have been studied...there seems to be some limitation built into us either by learning or by the design of our nervous systems, a limit... [on] our channel capacities...” (Miller, 1956, p.86)
To properly gain knowledge from instruction, students must:

1. Attend to presented material;
2. Mentally organize the material into a coherent structure; and
3. Integrate the material with existing knowledge.
Cognitive Load Theory

**Intrinsic Load**
Amount of mental processing requisite for completing a task.

**Germane Load**
Cognitive demand that is not necessary for gaining essential knowledge but enhances learning through automation or generalization.

**Extraneous Load**
Demand for cognitive resources to attend to and integrate nonessential elements that are preliminary to actual learning.

“Intrinsic, extraneous, and germaine cognitive loads are additive in that, together, the total load cannot exceed the working memory resources available if learning is to occur” (Paas, Renkl, and Sweller, 2003, p.2).
Cognitive Load Theory

- **Split attention effect** (Chandler & Sweller, 1996)
- **Redundancy effect** (Mayer & Moreno, 2003)
- **Seductive detail effect** (Graves et al., 1988; 1991; Wade, Schraw, Buxton, & Hayes, 1993; Goetz & Sadoski, 1995)

(Mayer & Moreno, 2003)
Downing, Haladyna, and Rodriguez (2002) synthesized test & item-writing research into 31 guidelines for writing multiple-choice items.

Content concerns
1. Every item should reflect specific content and a single specific mental behavior, as called for in text specifications (two-way grid, test blueprint).
2. Base each item on important content to learn; avoid trivial content.
3. Use novel material to test higher level learning. Paraphrase textbook language or language used during instruction when used in a test item to avoid testing for simply recall.
4. Keep the content of each item independent from content of other items on the test.
5. Avoid over specific and over general content when writing MC items.
6. Avoid opinion-based items.
7. Avoid trick items.
8. Keep vocabulary simple for the group of students being tested.

Formatting concerns
9. Use the question, completion, and best answer versions of the conventional MC, the alternate choice, true-false (TF), multiple true-false (MTF), matching, and the context-dependent item and item set formats, but AVOID the complex MC (Type K) format.
10. Format the item vertically instead of horizontally.

Style concerns
11. Edit and proof items.
12. Use correct grammar, punctuation, capitalization, and spelling.
13. Minimize the amount of reading in each item.

Writing the stem
14. Ensure that the directions in the stem are very clear.
15. Include the central idea in the stem instead of the choices.
16. Avoid window dressing (excessive verbiage).
17. Word the stem positively, avoid negatives such as NOT or EXCEPT. If negative words are used, use the word cautiously and always ensure that the word appears capitalized and boldface.

Writing the choices
18. Develop as many effective choices as you can, but research suggests three is adequate.
19. Make sure that only one of these choices is the right answer.
20. Vary the location of the right answer according to the number of choices.
21. Place choices in logical or numerical order.
22. Keep choices independent; choices should not be overlapping.
23. Keep choices homogeneous in content and grammatical structure.
24. Keep the length of choices about equal.
25. None-of-the-above should be used carefully.
26. Avoid All-of-the-above.
27. Phrase choices positively, avoid negatives such as NOT.
28. Avoid giving clues to the right answer, such as
   a. Specific determiners including always, never, completely, and absolutely.
   b. Clang associations, choices identical to or resembling words in the stem.
   c. Grammatical inconsistencies that cue the test-taker to the correct choice.
   d. Conspicuous correct choice.
   e. Pairs or triplets of options that clue the test-taker to the correct choice.
   f. Blatantly absurd, ridiculous options.
29. Make all distractors plausible.
30. Use typical errors of students to write your distractors.
31. Use humor if it is compatible with the teacher and the learning environment.
TAMI:
Content and Structure

\[ (a+b)^n = (a+b)^n \]

Beddow
Anatomy of an Item

Stimulus

Visual
(pictures, graphs, tables, charts, or figures)

Stem

Answer Choices
key (B) and distractors (A & C)

Page Layout

Mr. Murphy uses his car to get to work three days each week.

How many miles does Mr. Murphy drive to and from his job each week?

- 60 miles
- B. 120 miles
- C. 200 miles
The **TAMI** consists of two components:

1. The original **Inventory**:  
   a) Item Analysis (51 descriptors); and  
   b) Computer-Based Test Analysis (35 descriptors)

2. The **Accessibility Rating Matrix (ARM)**.
The TAMI Accessibility Rating Matrix (ARM) contains two rubrics.

- **Item Analysis** (i.e., item element) rubric:
  - Passage / Item Stimulus
  - Item Stem
  - Visuals
  - Answer Choices
  - Page / Item Layout

- **Overall Analysis** (i.e., overall item-level) rubric.

**ARM Accessibility Levels**

- **4** Maximally Accessible for Nearly All Test-Takers
- **3** Maximally Accessible for Most Test-Takers
- **2** Maximally Accessible for Some Test-Takers
- **1** Inaccessible for Many Test-Takers
3. Find $x$. 

Here it is.
# Item Analysis

<table>
<thead>
<tr>
<th><strong>Level</strong></th>
<th><strong>Inaccessible for Many Test-Takers</strong></th>
<th><strong>Maximally Accessible for Some Test-Takers</strong></th>
<th><strong>Maximally Accessible for Most Test-Takers</strong></th>
<th><strong>Maximally Accessible for Nearly All Test-Takers</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level 1</strong></td>
<td>Contains many words that are not essential for responding to the item(s).</td>
<td>Contains some words that are not essential for responding to the item(s).</td>
<td>Contains a few words that are not essential for responding to the item(s).</td>
<td>Contains only words that are essential for responding to the item(s).</td>
</tr>
<tr>
<td></td>
<td>The majority of text is likely to be difficult to understand for some test-takers.</td>
<td>A large portion of text is likely to be difficult to understand for test-takers.</td>
<td>Some text is likely to be difficult to understand for test-takers.</td>
<td>Text is minimal in length and written as plainly as possible.</td>
</tr>
<tr>
<td></td>
<td>Vocabulary and sentence structure are not grade-appropriate.</td>
<td>Vocabulary and sentence structure are mostly grade-appropriate.</td>
<td>Vocabulary and sentence structure are mostly grade-appropriate.</td>
<td>Vocabulary and sentence structure are grade-appropriate.</td>
</tr>
<tr>
<td></td>
<td>Idioms or jargon are used liberally.</td>
<td>Idioms or jargon are used frequently.</td>
<td>Idioms or jargon are used minimally.</td>
<td>Idioms or jargon are avoided.</td>
</tr>
<tr>
<td><strong>Level 2</strong></td>
<td>The entirety of the stem is overly complex.</td>
<td>Much of the stem language is overly complex.</td>
<td>Contains some text that could be simplified.</td>
<td>Text is minimal in length, written as plainly as possible.</td>
</tr>
<tr>
<td></td>
<td>Vocabulary and sentence structure are not grade-appropriate.</td>
<td>Vocabulary and sentence structure are mostly grade-appropriate.</td>
<td>Vocabulary and sentence structure are grade-appropriate.</td>
<td>Vocabulary and sentence structure are grade-appropriate.</td>
</tr>
<tr>
<td></td>
<td>Does not reflect intended content standard(s) and/or objective(s).</td>
<td>Reflects intended content standard(s) and/or objective(s).</td>
<td>Reflects intended content standard(s) and/or objective(s).</td>
<td>Reflects intended content standard(s) and/or objective(s).</td>
</tr>
<tr>
<td></td>
<td>Stem directive or question is very confusing.</td>
<td>Stem directive or question is somewhat confusing.</td>
<td>Target construct is evident.</td>
<td>Target construct is evident.</td>
</tr>
<tr>
<td></td>
<td>Uses not or except.</td>
<td>Uses not or except.</td>
<td>Positively worded, written in the active voice.</td>
<td>Positively worded, written in the active voice.</td>
</tr>
<tr>
<td></td>
<td>Written in the passive voice.</td>
<td>Written in the active voice.</td>
<td>Target construct is evident.</td>
<td>Target construct is evident.</td>
</tr>
</tbody>
</table>

- **Visuals**
  - Included visual(s) are irrelevant, unnecessary, and may cue the test-taker to an incorrect response.
  - Included visual(s) are necessary, but poorly depict the intended image(s).
  - Visuals contain a large amount of unnecessary complexity and text.
  - Visual(s) likely will cause confusion for test-takers, possibly cueing to an incorrect response.

- **Answer Choices**
  - Contains many nonessential words.
  - The complexity or confusing nature of answer choices is likely to cause problems for some test-takers.
  - Key and distractors are unbalanced with regard to order, length, or content in a manner that is likely to cue test-takers to an incorrect response.
  - One or more distractors is implausible.
  - More than one answer choice may be correct.

- **Page/Item Layout**
  - Page and/or item layout appears very cluttered and confusing; font sizes are too small.
  - A large amount of information is spread across multiple pages/screens.
  - Inessential page elements are distracting, draw attention from item elements that are necessary for responding.
  - Visuals are not integrated with the item stimulus and stem.
ARM Overall Analysis

Overall Analysis

4. Maximally Accessible for Nearly All Test-Takers
- Item contains only content (words, visuals) that is essential for responding to the item.
- All item text is minimal in length and written as plainly as possible.
- **Bold** font is used for key words.
- Item stem is positively worded, written in the active voice, and the target construct is evident.
- Any included visuals are necessary and clearly depict the intended image(s).
- All answer choices are necessary, plausible, and balanced with regard to length, content, and order.
- Entire item is presented together on one page/screen in a manner that facilitates responding.

3. Maximally Accessible for Most Test-Takers
- Item contains some content that is not essential for responding to the item.
- **Bold** font is used for key words.
- Stem is positively worded, written in the active voice, and the target construct is evident.
- Included visuals are not as simple or clear as possible.
- Visuals are not integrated with the other item elements.
- One or more distractors is unnecessary and/or answer choices are unnecessarily complex or unbalanced with regard to length, content, and order. Only one option is correct.
- Item layout is somewhat cluttered, or test-taker must turn the page to respond to the item.

2. Maximally Accessible for Some Test-Takers
- Item contains content that is not essential for responding to the item, to the extent that it may be distracting or confusing to the test-taker.
- The wording of the item stem may cause some confusion as to what is required.
- Included visuals are unnecessary and potential distract the test-takers from essential item elements, or visuals are do not clearly depict the intended images or are unnecessarily complex.
- One or more distractors is implausible or absurd. Only one option is correct.
- Nonessential item elements in the page layout may draw test-taker attention away from essential content, or the test-taker must turn the page 2 or more times to respond to the item.

1. Inaccessible for Many Test-Takers
- The item contains a large amount of content that is not essential for responding to the item, to the extent that it is likely to confuse the test-taker.
- Stem is negatively worded, in passive voice, and/or it is not evident what is required.
- Included visuals are irrelevant and may cue test-taker to an incorrect response, or included visuals are likely to confuse the test-taker due to complexity or lack of clarity.
- Answer choices are unbalanced in a manner that may cue an incorrect response, contain more than one correct answer, and/or are implausible/absurd.
- Nonessential item elements in the page layout are likely to draw attention from essential information, or a large amount of essential information is presented across multiple pages/screens.
TAMI:
Application
Mr. Jameson is a salesman who lives in Knoxville, Tennessee. He travels the distance from his home to Jackson and back twice a month. He also travels to Chattanooga and back home twice a month. If the distance from Knoxville to Jackson is 295 km and the distance from Knoxville to Chattanooga is 95 km, how many km does he drive on these trips each month?

- A. 670 km
- B. 780 km
- C. 1,560 km
- D. 14,103 km

<table>
<thead>
<tr>
<th>AA-MAS Eligibility</th>
<th>Response Frequency</th>
<th>Item Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Not eligible</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eligible</td>
<td>25%</td>
<td>36%</td>
</tr>
</tbody>
</table>

*key; R = Readability Score
5. Mr. Jamieson is a salesman who lives in Knoxville, Tennessee. He travels the distance from his home to Jackson and back twice a month. He also travels to Chattanooga and back home twice a month. If the distance from Knoxville to Jackson is 295 km and the distance from Knoxville to Chattanooga is 95 km, how many km does he drive on these trips each month?

- A. 670 km
- B. 780 km
- C. 1,560 km
- D. 14,103 km

<table>
<thead>
<tr>
<th>AA-MAS</th>
<th>Eligibility</th>
<th>Response Frequency</th>
<th>Item Information</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not eligible</td>
<td>16% 50% 30% 4%</td>
<td>p 0.30 DOK 2 R 8.7 # Words 66</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Eligible</td>
<td>25% 36% 21% 15%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*key; R = Readability Score

**Overall Rating:**
Maximally Accessible for Some Test-Takers
5. Mr. Jameson is a salesman who lives in Knoxville, Tennessee. He travels the distance from his home to Jackson and back twice a month. He also travels to Chattanooga and back home twice a month. If the distance from Knoxville to Jackson is 295 km and the distance from Knoxville to Chattanooga is 95 km, how many km does he drive on these trips each month?

- A. 670 km
- B. 780 km
- C. 1,560 km
- D. 14,103 km

- Reduced extraneous verbiage in stimulus and stem.
- Added **bold** to key words.
- Added a visual.
- Reduced the number of distractors.
- Increased white space.
### Original Item

<table>
<thead>
<tr>
<th>AA-MAS Eligibility</th>
<th>Response Frequency</th>
<th>Item Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Not eligible</td>
<td>16%</td>
<td>50%</td>
</tr>
<tr>
<td>Eligible</td>
<td>25%</td>
<td>36%</td>
</tr>
</tbody>
</table>

### Modified Item

<table>
<thead>
<tr>
<th>AA-MAS Eligibility</th>
<th>Response Frequency</th>
<th>Item Information</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>Not eligible</td>
<td>16%</td>
<td>45%</td>
</tr>
<tr>
<td>Eligible</td>
<td>29%</td>
<td>36%</td>
</tr>
</tbody>
</table>

*key; R = Readability Score

### Summary

- **DOK:** No change
- **Readability:** -3.1
- **# of Words:** -9 (↓ 14%)
Tests can be improved for students with and without special needs!
Thank you.

peter.beddow@vanderbilt.edu

http://peabody.vanderbilt.edu/tami.xml


