



Investigation of Some Assessment Issues for Students who are Blind or Visually Impaired

Elizabeth Stone
Educational Testing Service

NERA Annual Conference
October 18, 2007

Overview

- Goals of TARA™ project
- DIF/DDF Study results
- Test form review
- Plans for future research

TARA™

The Technology Assisted Reading Assessment (TARA™) project has two main purposes:

- To research and develop standards-based reading assessments for students who are blind or visually impaired
- To research the design and development of an alternate assessment of modified achievement standards in reading for examinees who are blind or visually impaired

The current study is part of the research phase of the project.

Nature and Purpose of DIF and DDF Studies

- Evaluate comparability of measurement characteristics for students without disabilities and students who are blind or visually impaired
 - No reason to consider overall abilities of groups to be different
 - Look for construct-irrelevant causes
- A priori DIF hypotheses (position, visual content, textual content, language, format)

Description of Test

- ELA component of large-scale state standards test
- 75 multiple choice questions
 - 56% Reading / 44% Writing
 - Essay component in 4th grade, excluded from analyses

Description of Samples: Summary Statistics

| Grade | Summary Statistics | Students without Disabilities (sample; non-ELL only) | Students who are Blind or Visually Impaired (all; English Language Learner + non-ELL) | |
|-------|--------------------|--|---|------------------------|
| | | | Large Print | Large Print or Braille |
| | | Reference | Focal | Focal |
| 4 | N | 30225 | 105 | 141 |
| | Mean | 48 | 37 | 37 |
| | SD | 14 | 15 | 15 |

Methods

- DIF (Differential Item Functioning): See if groups seem to get item right or wrong in different proportions after being matched on ability
 - Mantel-Haenszel with purification
 - ELA score as matching criterion
 - ETS delta-DIF categories (-C, -B, A, B, C)
- DDF (Differential Distractor Functioning): See if groups seem to choose distractors in different proportions after being matched on ability
 - Standardization method
 - Typical cut-offs for significance

DIF Results

Displayed B or C DIF

Hypothesized
to show DIF

| | Yes | No | |
|-----|----------|-----------|-----------|
| Yes | 3 | 41 | 44 |
| No | 6 | 25 | 31 |
| | 9 | 66 | 75 |

DIF Results

- A priori hypotheses were not supported
- Number of DIF items smaller than expected (sample size issue?)
- Reading/Writing divide?
 - Most Reading items displaying DIF favored the focal group (4/5)
 - Most Writing items displaying DIF favored the reference group (4/5)
- Effects did not seem to apply for all items associated with a passage

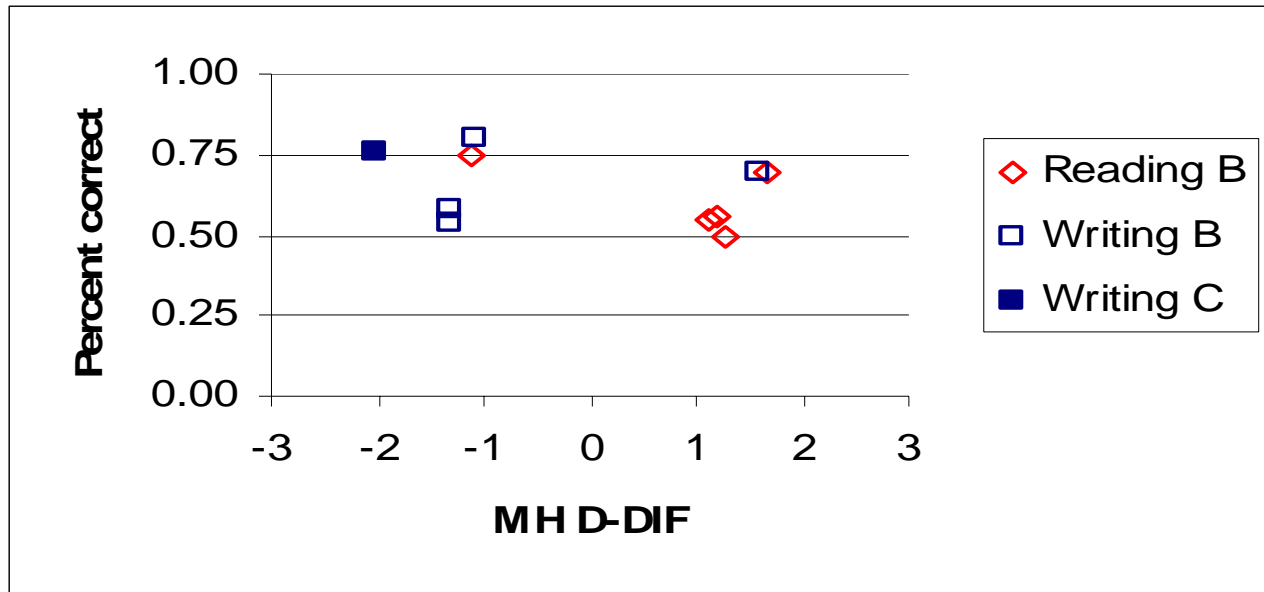
DIF Results

- Where DIF did seem to show up:
 - Metaphor
 - Items and passages related to how people experience the world (e.g. through touch)
 - Unusual document format
- Where DIF did not seem to show up:
 - Items and passages involving typically “sighted” activities or interests, e.g. photography (in fact, some favor focal groups)

Reference: Students without Disabilities Taking Standard Form

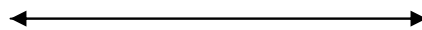
Focal: Students Who are Visually Impaired Taking Large Print Form

Easy



Difficult

Favors Students without
Disabilities

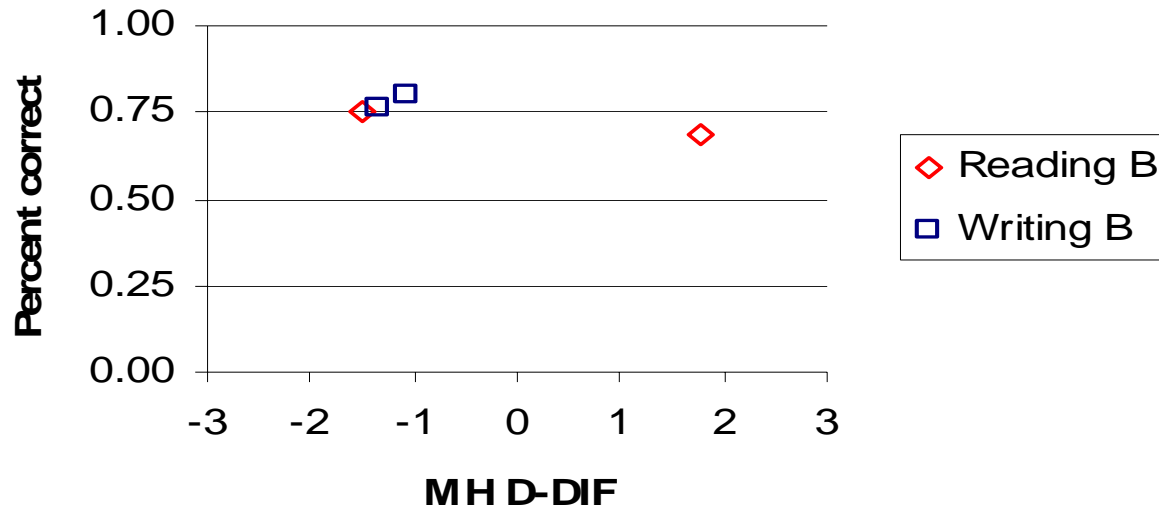


Favors Students who
are Blind or Visually
Impaired Using Large
Print

Reference: Students without Disabilities Taking Standard Form

Focal: Students Who are Blind or Visually Impaired Taking Large Print or Braille Form

Easy



Difficult

Favors Students without Disabilities



Favors Students who are Blind or Visually Impaired Using Large Print or Braille

DDF Results

Results were not as interpretable as hoped

- For the large print focal group (10 B or C DIF items): 2 items had a highly significant distractor; 8 items had at least one moderately significant distractor
- For the large print or braille focal group (5 B or C DIF items): 1 item had a highly significant distractor; 3 items had at least one moderately significant distractor
- Review did not reveal any obvious causes

Reference: Students without Disabilities Taking Standard Form

Focal: Students who are Visually Impaired Taking Large Print Form

| item | Reference P+ | Focal P+ | MH D-DIF | Category | Standardized p-DIF | A | B | C | D | no response |
|------|--------------|----------|----------|----------|--------------------|--------------|--------------|--------------|-------------|-------------|
| 21 | 0.55 | 0.50 | 1.22 | B + | 0.10 | -0.11 | -0.01 | 0.01 | 0.10 | 0.01 |
| 23 | 0.68 | 0.65 | 1.70 | B + | 0.13 | -0.05 | 0.15 | -0.12 | 0.02 | 0.01 |
| 24 | 0.51 | 0.50 | 1.28 | B + | 0.12 | 0.12 | -0.07 | -0.01 | -0.04 | 0.01 |
| 26 | 0.55 | 0.47 | 1.06 | B + | 0.09 | -0.02 | -0.06 | 0.09 | -0.03 | 0.01 |
| 32 | 0.75 | 0.50 | -1.15 | B - | -0.10 | 0.04 | 0.03 | -0.11 | 0.03 | 0.01 |
| 56 | 0.79 | 0.59 | -1.14 | B - | -0.09 | -0.11 | 0.03 | 0.02 | 0.06 | 0.00 |
| 58 | 0.58 | 0.32 | -1.25 | B - | -0.10 | 0.07 | 0.03 | 0.01 | -0.10 | 0.00 |
| 67 | 0.76 | 0.42 | -2.11 | C - | -0.18 | -0.17 | 0.05 | 0.06 | 0.05 | 0.01 |
| 71 | 0.53 | 0.27 | -1.31 | B - | -0.09 | 0.02 | -0.03 | 0.07 | -0.09 | 0.03 |
| 75 | 0.70 | 0.68 | 1.46 | B + | 0.13 | -0.08 | -0.04 | 0.12 | -0.03 | 0.03 |

Reference: Students without Disabilities Taking Standard Form

Focal: Students who are Blind or Visually Impaired Taking Large Print or Braille Form

| item | Reference P+ | Focal P+ | MH D-DIF | Category | Standardized p-DIF | A | B | C | D | no response |
|------|--------------|----------|----------|----------|--------------------|--------------|-------|--------------|-------------|-------------|
| 21 | 0.55 | 0.50 | 1.03 | B + | 0.09 | -0.07 | -0.03 | 0.01 | 0.09 | 0.01 |
| 23 | 0.68 | 0.64 | 1.74 | B + | 0.12 | -0.02 | 0.12 | -0.12 | 0.01 | 0.01 |
| 32 | 0.75 | 0.48 | -1.60 | B - | -0.13 | 0.05 | 0.03 | -0.14 | 0.04 | 0.02 |
| 56 | 0.79 | 0.60 | -1.16 | B - | -0.10 | -0.10 | 0.03 | 0.02 | 0.06 | 0.00 |
| 67 | 0.76 | 0.48 | -1.43 | B - | -0.11 | -0.11 | 0.01 | 0.04 | 0.04 | 0.02 |

Review of Forms

Large print form issues:

- Portions were merely an enlargement, so some fonts, e.g. footnotes, were not in proper font size.
- As with the standard form, many different fonts were used. This is not considered “best practice” for large print.
- We noted one passage description that was slightly different, and one item that had different wording from the standard form.

Review of Forms

Braille form issues:

- Distracting logo encoding and page numbering
- Paragraph numbering
- Contracted vs. uncontracted
- Symbols used
 - Symbols not encountered before (e.g., italics)
 - Nemeth code vs. literary braille

Discussion

There are some caveats:

- Sample sizes smaller than desired: These are “low incidence” populations.
- We would need more of each type of item to support inference of a pattern.

However, some inferences can be drawn that may be useful in instruction, test development, administration, and research:

- access to curriculum
- access to various document formats for test preparation
- careful review of test forms for these issues
- importance of proctor awareness
- need to include these groups in fairness and validity measures

Future Steps for TARA™

- Examine items and item types across administrations
- Investigate other DIF methods
- Test analysis
- Trend analysis
- Contribute test development recommendations
- Creation of TARA™ using ECD

Please feel free to email me at

ESTONE@ETS.ORG

with any questions, comments, or requests for papers.

Thank you!

