

Appendix G

Data Review Training Slides

AzMERIT | Arizona's Statewide Achievement Assessment for English Language Arts and Mathematics

Statistical Review Training for ADE

AIR

AzMERIT | Arizona's Statewide Achievement Assessment for English Language Arts and Mathematics

Statistical Review of Items

- Item Quality and Performance
 - Does the item behave the way it's supposed to behave?
- Item Difficulty
 - How hard is the item?
- Differential Item Functioning
 - Does the item behave differently across subgroups?

AIR 2

AzMERIT | Arizona's Statewide Achievement Assessment for English Language Arts and Mathematics

Item Quality

- Do highly skilled students perform better on the item than less skilled students?
- Correlation with Test – link between selecting a response option and doing well on the rest of the test
 - For key, + is good, – is bad
 - For distracters, – is good, + is bad

AIR 3

AzMERIT | Arizona's Statewide Achievement Assessment for English Language Arts and Mathematics

Item Quality Flag Criteria

- Adjusted biserial/polyserial correlation statistic is less than .25 for multiple-choice or constructed-response items; (AB)
- Adjusted biserial correlations for multiple-choice item distractors is greater than .05; (ABD)

AIR 4

AzMERIT | Arizona's Statewide Achievement Assessment
for English Language Arts and Mathematics

Item Difficulty

- How hard is the item?
- What percent of students answer item correctly?
- MC items – % of students selecting each response option
- Non-MC items – % of students achieving each score point

AIR | ARIZONA INSTITUTE FOR RESEARCH

5

AzMERIT | Arizona's Statewide Achievement Assessment
for English Language Arts and Mathematics

Item Difficulty Flag Criteria

- Proportion correct value is less than .25 or greater than .95 for multiple-choice items, or greater than .95 for any single score point of a constructed-response item;
- Also known as p-value (P or CR_Prop)

AIR | ARIZONA INSTITUTE FOR RESEARCH

6

AzMERIT | Arizona's Statewide Achievement Assessment
for English Language Arts and Mathematics

Non-Modal Key

- A distractor is chosen by students more often than the key is chosen

AIR | ARIZONA INSTITUTE FOR RESEARCH

7

AzMERIT | Arizona's Statewide Achievement Assessment
for English Language Arts and Mathematics

Non-Modal Key Flag Criteria

- The proportion of students responding to a distractor exceeds the proportion responding to the keyed response for MC items; (NMK)

AIR | ARIZONA INSTITUTE FOR RESEARCH

8

AzMERIT | Arizona's Statewide Achievement Assessment for English Language Arts and Mathematics

Omit Rate

- Students do not provide a response

AIR | 9

AzMERIT | Arizona's Statewide Achievement Assessment for English Language Arts and Mathematics

Omit Rate Flag Criteria

- Omit rate is greater than .15;

AIR | 10

AzMERIT | Arizona's Statewide Achievement Assessment for English Language Arts and Mathematics

Differential Item Functioning

- * Fair Items behave similarly across groups
- * Probability of answering correctly is the same for all students of similar ability regardless of group membership

Subgroup Comparisons:

- Female/Male
- Non-Hispanic / Hispanic, Latino or Spanish origin
- Black, African American / White
- American Indian or Alaskan Native / White
- Asian / White
- Native Hawaiian or Other Pacific Islander / White
- Multiple ethnicities selected / White

AIR | 11

AzMERIT | Arizona's Statewide Achievement Assessment for English Language Arts and Mathematics

Differential Item Functioning (DIF)


- Direction of possible bias
 - “-” item favors reference groups
 - “+” item favors focal group
- Severity of possible bias
 - “A” No statistical evidence of DIF
 - “B” Evidence for potential mild DIF
 - “C” Evidence for potential severe DIF
- “C” indicates that the item is more difficult for one group and should be reviewed carefully for bias

AIR | 12

AzMERIT | Arizona's Statewide Achievement Assessment
for English Language Arts and Mathematics

DIF Flag Criteria

- Items are classified into three categories (A, B, or C), ranging from no evidence of DIF to severe DIF.
- Items are categorized as **positive DIF** (i.e., +A, +B, or +C), signifying that the item **favors the focal group** (e.g., African American/Black, Hispanic, or female), or
- **negative DIF** (i.e., -A, -B, or -C), signifying that the item **favors the reference group** (e.g., white or male).
- Items are flagged if their DIF statistics fall into the “C” category for any group, which indicates that the item shows **significant DIF** and should be reviewed for potential content bias, differential validity, or other issues that may reduce item fairness


 AIR | ARIZONA INSTITUTE FOR RESEARCH

13

AzMERIT | Arizona's Statewide Achievement Assessment
for English Language Arts and Mathematics

Content Expert Judgments

- Statistical information is important, but not a substitute for expert judges
- Items central to a learning standard may be difficult because a concept is not currently included in curriculum
- Items may show DIF because some concepts may be less likely to be covered in all area schools


 AIR | ARIZONA INSTITUTE FOR RESEARCH

14

AzMERIT | Arizona's Statewide Achievement Assessment
for English Language Arts and Mathematics

Logistics

- Items can be found at the **Content and Fairness Data Review and Resolution** review level in the Arizona Assessment project in ITS
- The MDSs will be posted here on the sftp:
[/files/AzMERIT/To ADE/Content Data Review/](#)
- Please “PEND” any data comments in ITS

 AIR | ARIZONA INSTITUTE FOR RESEARCH

15