

Maryland's Test Development Process

Report to the State Board of
Education

December 18, 2008

How Maryland Tests Are Developed

Assessment Documents

VSC (MSA)

- G 3-8
- Reading
- Math
- Science

CLG (HSA)

- Biology
- English
- Government
- Algebra/
Data Analysis

Assessment Limits

- Skills, concepts, and processes that are “fair game” for testing
- Identified by MD educators
- Part of public assessment documents (VSC, CLG)

Test Maps

Blueprints

- Specify extent of skill/concept measurement
- Specify test make up
- Take sampling into account
- Aligned to Assessment Documents and Limits

Test Construction

- Requirements of test maps/blueprints
 - CRT
 - Item types
- Prior test data
- Timing
- Embedded field-test items

Anatomy of the VSC

Standard: *Knowledge of number relationships and computation/arithmetic*

Topic: *Number computation*

Indicator: *Analyze number relations and compute*

Objective: *Divide whole numbers*

Assessment Limit: *Use up to a 3-digit dividend by a 1-digit divisor and whole numbers with no remainders (0 - 999)*

Sample of a Test Blueprint

Reporting Category	Item Type			Total Items
	SR	BCR	ECR	
Total Test	53	6	1	60
Skills/Processes	7	2	0	9
Earth/Space	10	1	0	11
Life	9	1	0	10
Chemistry	10	0	0	10
Physics	8	1	0	9
Environment	9	1	1	11

Item & Test Development

Writing & Editing

Item Writers

- Trained content specialists
- Assignments based on blueprints and needs
- Quality checks
- Item specifications

Editors

- Content
- Style

Reviews

- Side-by-side with MSDE specialists
- Content Reviews with MD educators
- Bias and Sensitivity Reviews for diverse populations

Page Production

- Manuscript
- 1st pages/ lasers
- 2nd pages/ lasers
- Final

Delivery

- To schools and school systems

Field Testing

Administration

Initial Admin

- Census
- Random distribution of forms

On-going

- Embedded in operational test

Reviews

- Item statistics
- DIF Analyses

Item Selection

- Entered into bank
- Re-written and re-field tested
- Thrown out

Operational Test Forms

Construction

- Banked items
- Field test items

DIF Example

Concept: The odds of answering the item correctly should be the same regardless of group

- Define a “reference” and “focal group” (male/female, African American/White)
- Calculate odds ratio: odds of correct response by reference group divided by the odds for the focus group

DIF Example

Null hypothesis: Odds ratio = 1 (odds of getting the item correct are equal for the two groups)

Odds of answering correctly:

Males 70%

Females 50%

DIF = 2.33 favoring males

Scoring Constructed Response Items

Tools

- Good Items
- Rubrics
 - Generic rubrics
- Anchor Papers
 - Same for each admin. and item type
 - Prevent drift

Range Finding

- Post field test
- MSDE content and scoring specialists
- Maryland educators
- Read, discuss, identify papers for scoring guides

Scoring Guides

- Item specific
- Student responses representing range of scores
- Annotated according to criterion of rubrics

Quality Control

- Training
- Anchor papers
 - Item-specific
 - Anchor Sets
 - Scoring Guides
 - Check sets
 - Read-Behinds
 - Retraining
 - Resolution

Sample Rubric – MSA Reading BCR

Score 3

The response demonstrates an understanding of the complexities of the text.

Addresses the demands of the question

Effectively uses text-relevant information to clarify or extend understanding

Score 2

The response demonstrates a general understanding of the text.

Partially addresses the demands of the question

Uses text-relevant information to show understanding

Score 1

The response demonstrates a minimal understanding of the text.

Minimally addresses the demands of the question

Uses minimal information to show some understanding of the text in relation to the question

Score 0

The response is completely incorrect, irrelevant to the question, or missing.

Scoring Guide Paper

Back then you would focus on just one subject such as being a blacksmith and then when you grew up, chances are you would end up being a blacksmith. Now, today you learn a lot of different things and you don't decide your profession until you are done with your education. Nathan knows he will probably end up being a blacksmith, but we have a long time before we decide what we'll do for a living.

Anchor Paper Score 2

This response indicates a good understanding of the text. The student identifies a comparison, “Back then your would focus on one subject....now, today you learn a lot of different things” the students uses text-relevant information to explain the comparison, “Nathan knows he will probably end up being a blacksmith but we have a long time before we decide what we'll do for a living.”

Standard Setting

Who

- Content Experts
- Psychometricians
- Stakeholders

Process

- Training
- Ordered item booklets
- Bookmarking procedure
- Rounds and discussion

Approval

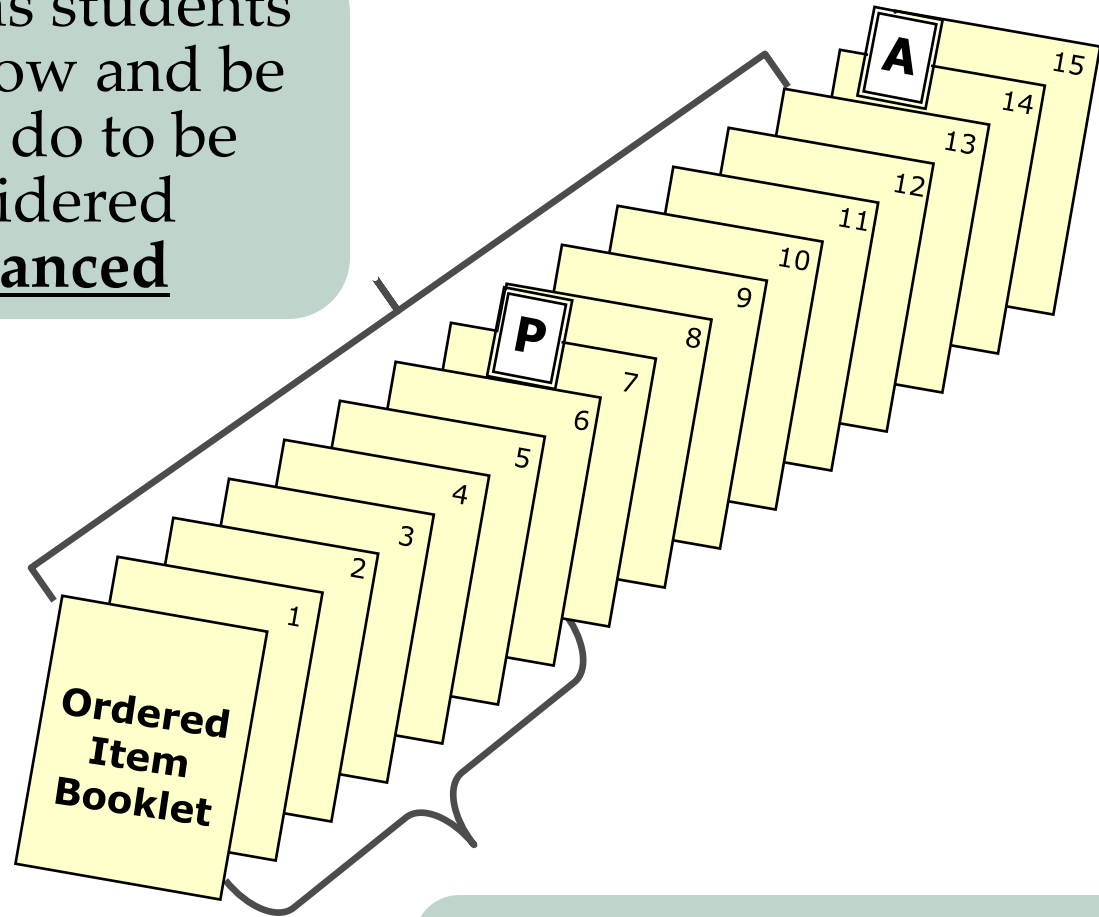
- Certification of process
- Impact data and verification
- State Board of Education

Final Scoring

- Standard applied to operational items
- Scale scores established
- Scores generated and reported

Bookmark Standard-Setting Procedure

Test items students must know and be able to do to be considered Advanced



Test items students must know and be able to do to be considered Proficient

Equating

Purpose

Maintain Standard

- Year to year
- Test to test

Process

- Core set of items (anchors)
- Use original item statistics
- Statistical procedure

Adjustments

- Test difficulty
- Test Characteristic Curve
- Maintain passing scale score

Result

- Same standard and difficulty
- Scale scores consistent
- Fair and objective

Reporting

Step 1

Validate data files for processing

- Grade
- Gender
- Race
- Services

Step 2

- Complete preliminary reports
- Send to school systems

Step 3

- Complete appeals process for AYP
- Send final reports to school systems

Step 4

- Report results to Board of Education
- Post results on website

More Information

- www.MdReportCard.org
- www.HSAexam.org
- www.MdK12.org