GEORGIA DEPARTMENT OF EDUCATION

Kathy Cox
State Superintendent of Schools

WHO ARE THOSE 2% STUDENTS AND HOW DO WE DESIGN ITEMS THAT WORK?

CCSSO NATIONAL CONFERENCE ON STUDENT ASSESSMENT
JUNE 2008 / ORLANDO, FL

Presenters

- Melissa Fincher, Georgia Department of Education
- Jenn Dunn, Measured Progress
- Patti McDivitt, DRC
- Karin Hess, NCIEA

- Discussant: Sue Rigney, US ED

Georgia's EAG

- The work described in this session was done through an EAG awarded to Georgia in October 2007.
  - Georgia partnered with Hawaii and Kentucky to run separate but related studies
  - Georgia's partners include:
    - National Center for Educational Outcomes
    - Southeastern Regional Resource Center
    - National Center for Improving Educational Assessments
    - Patti McDivitt
    - Georgia Center for Assessment (UGA)

Georgia's EAG Goals

1. Understand who the lowest performing students are
2. Understand the achievement of these students
3. Evaluate Georgia's current assessments in light of what is learned about these students
4. Pilot and evaluate additional access methods

Why this approach?

- Georgia believes that in order to build a technically sound assessment, it is imperative that we understand who the target population is and what they can do
  - Our goal is to target the assessment toward their strengths and provide scaffolding for their weaknesses

Context for Georgia's EAG Work

- The integrity of the assessment process must be protected

- Expectations for student achievement must be preserved
  - They affect opportunity to learn

Identification of the Population

- Persistently low performing
  - Achieve lowest performance level in Reading and/or Mathematics on three separate testing occasions
- 2006 Baseline Population
  - 2 Cohorts: Grade 5 and Grade 8
  - Assessment administrations:
    - Spring 2006
    - Spring 2005
    - Spring 2004

Center for Assessment

Demographic Trends

- More male students
  - Baseline: ~51% Identified: ~ 60% – 65%
- More black students
  - Baseline: ~40% Identified: ~ 60% – 65%
- More Free/Reduced Lunch students
  - Baseline: ~50% Identified: ~ 75% – 80%
- More students with disabilities
  - Baseline: ~15% Identified: ~ 40% – 55%
- More students with mild intellectual disabilities
  - Baseline: ~10% Identified: ~ 20% – 30%

Demographic Trends

- A higher proportion of ELL students are persistently low performing in reading (compared to the baseline).
- Approximately 2% of students at each grade were identified as persistently low performing in both content areas (Reading and Mathematics).
- In grade 8, 77% of the students identified in Reading were also identified in Mathematics.

Overall Process

- Independent qualitative and quantitative reviews of content standards, test blueprints, item specifications, student data, and existing test items
- Items identified as candidates for revision and/or enhancement
- Committee review by Georgia educators: items revised and/or enhanced without changing the intended construct
- Revised test blueprints (order/organization of items; revised & non-revised items in pilot test forms)
- Results of piloted items and curriculum implementation surveys analyzed

Quantitative Review

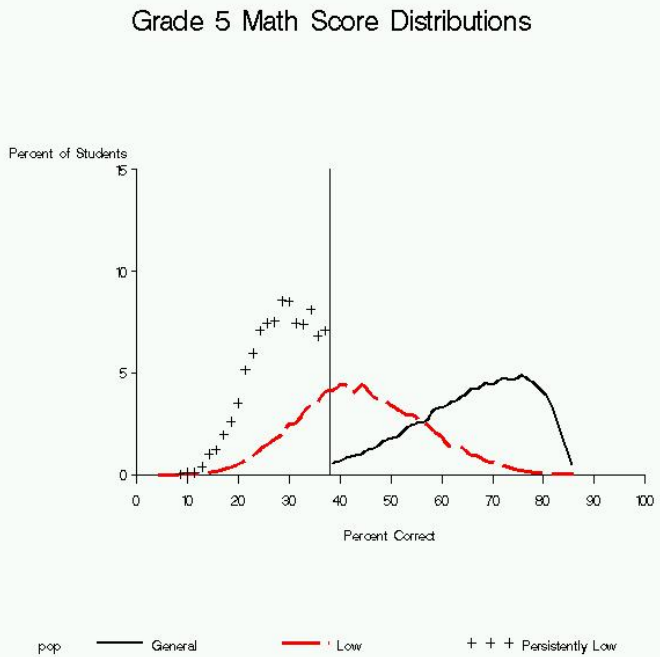
- Premise
  - A group of students that is persistently unable to meet the proficiency requirements exists
  - These students are able to demonstrate their knowledge and skills on some of the general assessment items but not on others
- Task
  - Develop a process to
    - Identify items that are “working” for these students
    - Identify items that are NOT “working” for these students
  - Investigate the skills and knowledge of persistently low performing students

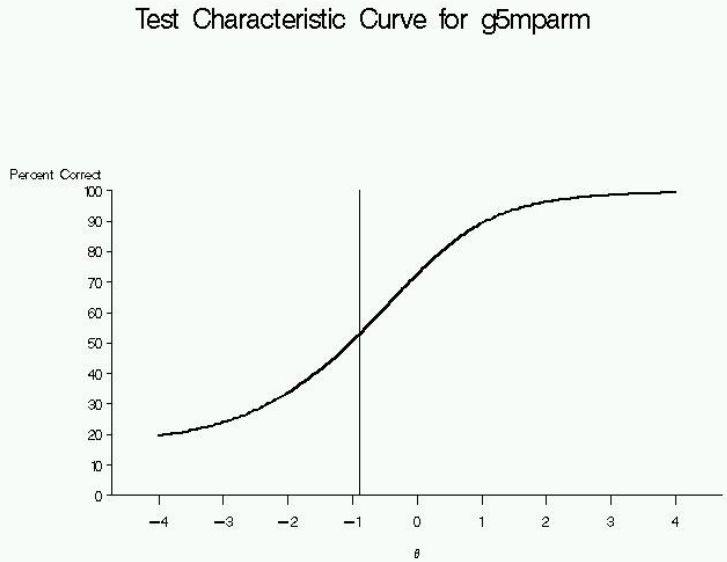
Identifying Knowledge and Skills

- Are there things that the persistently low students know?
  - Can we distinguish these students from low performing students?
- How to identify these skills?
  - Match the test to the population
  - Examine the item characteristics of the items that align with the population

Population Distributions and Test Location

- Where is the test in relation to each population:
  - Persistently Low
  - Low
  - General
    - Minus Low and Persistently Low





Identification of Items with Potential

1. Identify below the proficiency cut that meet IRT criteria (PLP)
   - Positive slopes
   - Difficulty parameters between -4 and +4
2. Identify items that meet classical criteria
   - Most PLP students selected the correct option
   - The correct option has the greatest discrimination
3. Identify items in Steps 1 & 2 that meet IRT criteria (PLP)
   - Positive slopes
   - Difficulty parameters between -4 and +4

	Number of Items	Step 1	Step 2	Step 3
Math: Grade 5	60	12	22	30
Math: Grade 8	60	26	16	38
Reading: Grade 5	40	9	14	14
Reading: Grade 8	40	12	6	15

Types of Knowledge and Skills

	Grade 5		Grade 8	
	Total Test	Effective Items	Total Test	Effective Items
Computation & Estimation	21%	33%	10%	11%
Geometry & Measurement	17%	20%	20%	21%
Number Sense & Numeration	20%	23%	14%	13%
Patterns & Relationships/Algebra	11%	13%	20%	26%
Problem Solving	20%	7%	20%	11%
Statistics & Probability	10%	3%	16%	18%
Total Number of Items	60	30	60	38

Types of Knowledge and Skills

	Grade 5		Grade 8	
	Total Test	Effective Items	Total Test	Effective Items
Functional & Media Literacy	15%	7%	18%	13%
Information	30%	43%	43%	40%
Literacy Comprehension	35%	25%	25%	27%
Skills and Vocabulary	20%	14%	15%	20%
Total Number of Items	40	14	40	15

Identification of Problematic Items

1. Identify items that DO NOT meet classical criteria
   - More than 35% of the PLP students select a particular incorrect option

2. Identify items that DO NOT meet classical criteria
   - A particular incorrect option has the greatest discrimination

3. Identify items below the cut score that do NOT meet IRT criteria

	Number of Items	Step 1	Step 2	Step 3	Final
Math: Grade 5	60	17	3	11	25
Math: Grade 8	60	16	6	9	21
Reading: Grade 5	40	8	0	20	22
Reading: Grade 8	40	1	0	13	14

Types of Knowledge and Skills

	Grade 5		Grade 8	
	Total Test	Ineffective Items	Total Test	Ineffective Items
Computation & Estimation	21%	8%	10%	10%
Geometry & Measurement	17%	20%	20%	14%
Number Sense & Numeration	20%	20%	14%	19%
Patterns & Relationships/Algebra	11%	12%	20%	14%
Problem Solving	20%	24%	20%	38%
Statistics & Probability	10%	16%	16%	5%
Total Number of Items	60	25	60	21

Types of Knowledge and Skills

	Grade 5		Grade 8	
	Total Test	Ineffective Items	Total Test	Ineffective Items
Functional & Media Literacy	15%	14%	18%	29%
Information	30%	23%	43%	43%
Literacy Comprehension	35%	36%	25%	21%
Skills and Vocabulary	20%	27%	15%	7%
Total Number of Items	40	22	40	14

What have we learned?

- Population
  - These students do have some grade-level knowledge and skills
  - Items exist for identifying these skills
    - Tend to be the easier items
- Statistical Techniques
  - Can be used to "cautiously" explore low ability populations
    - Cognizant of the context in which they are being used
    - Assumption of normal distributions
  - Looking at non-significant but possibly meaningful differences

Qualitative Reviews

- Independent content expert reviews of reading and mathematics items to select "candidate items" *in conjunction with* quantitative review
- Georgia educator reviews of selected candidate items and recommended revisions and/or enhancements
- Georgia DOE reviews of curriculum implementation survey and pilot test data for revised items

Questions Guiding the Qualitative Review of the Items [1]

- What skills and concepts are being tested?
- What is the focus of each item: application of skills, fact-based information, conceptual understanding, problem solving *AND what might make this difficult?*
- What is the vocabulary load of each item and the overall readability and text "density" for each item?
- Do the charts, tables, graphs, artwork, visuals, spacing used support *OR perhaps detract* from understanding?
- What is the "closeness" of distracters to each other?

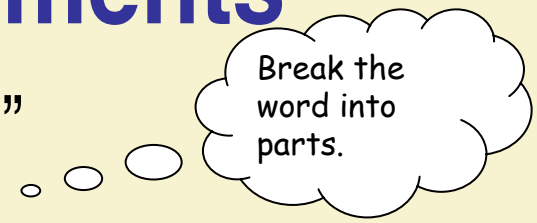
Questions Guiding the Qualitative Review of the Items [2]

- What is the complexity or abstractness of ideas presented (e.g., use of figurative language vs. literal meanings; theme vs. main idea)?
- For the mathematics items, it was also noted whether they were single-step or multi-step problems, required "extensive" reading, etc.
- For reading items, the genre, text features, text structure, readability, and length of each reading passage were also examined.

Suggested Revisions

- Simplify language in question/stem and/or distracters
- Simplify graphics, visuals, etc.
- Eliminate extraneous information
- Substitute another (*more familiar)* word without changing the construct
- Reformat items or passages (e.g., adding more white space, size of text)
- Adjust layouts (e.g., reorder items or passages)
- Separate reading passages into chunks, followed by related items

Suggested Enhancements



- Add a Helpful Hint in a "Thought Balloon"
  - Definition, key word or phrase
  - Reminder of approach to help solve a multi-step problem (e.g., circle the information you need to solve this)
- Provide a Scaffold
  - Graphic organizer (e.g., timeline for organizing chronology)
  - Table, graph, chart, or visual to enhance *conceptual understanding* (e.g., input-output chart for finding pattern)
  - Key words, phrases, definitions (e.g., sidebar glossary), introductions to passages
- General Presentation: underline, bold, enlarge key words/phrases/symbols; number paragraphs or lines

What Could Not Be Considered

- Shortening passages
- Adapting passages in any way that would make them "below grade level"
- Reducing the number of items per passage or adding more items to fewer passages
- Simply removing a distractor
- Revising questions or answer options in such a way as to change the nature of the item and/or the construct

Georgia Educator Review

- Reviewed the purpose of the assessment
- Defined/described the population
  - Who are these students? How would you describe them when they read or do mathematics in your classroom?
  - What are their greatest struggles in reading or mathematics?
  - What instructional strategies seem to best support their learning?
- Reviewed “candidate” items & item characteristics
- Suggested revisions and/or enhancements for each identified item, including visuals, graphics, etc.
- Provided suggestions for future item and test development

Qualitative Results: Reading [1]

Text Passages

- Separated passages into meaningful chunks, with accompanying items (field tested chunking in grade 8 only)
- Added directions to focus reading/set the purpose
- Added artwork or a graphic organizer
- Added more space between paragraphs

Qualitative Results: Reading [2]

- Revised Items by
  - Eliminating extraneous information
  - Simplifying the vocabulary
- Enhanced items by
  - Adding a graphic organizer
  - Adding a hint or thought balloon (e.g., Hint: What is the setting of the story?)
  - Enlarging the font
  - Reordering items “logically” (e.g., to follow passage sequence, analysis items after initial understanding)
  - Underlining key words and phases

Sample Training Example for Reading

(using more familiar words in distracters; shortening stem)

BEFORE

In paragraph 9, what does the word genuine mean?

a. real
b. pretend
c. content
d. anxious

AFTER

In paragraph 9, what does genuine mean?

a. real
b. pretend
c. happy
d. worried

Example of "Chunking" a passage

Carol Ryrie Brink lived with her Grandmother Caddie and two aunts on a tiny farm in Idaho. The farm was a great place to live. Besides wonderful tall climbing trees, it had cats and chickens and a barn that held her pony, Tommy, whom she rode all over the countryside and through the streets of town.

Even so, Carol was sometimes lonely. She would climb to the top of her favorite tree and make up stories to entertain herself. Carol knew all about stories. The favorite part of her day was when Grandmother Caddie would sit with her and tell stories about when she was a pioneer girl, and her family had moved from Boston to the wilderness of Wisconsin. Grandmother Caddie told Carol story after story about the adventures of her and her brothers.

Answer questions 1–6 in your answer booklet.

1. Why did Carol enjoy hearing her grandmother's stories?

   A She could listen to her grandmother's stories from the top of a tree.

   B She enjoyed how her grandmother read stories aloud.

   C She thought her grandmother's adventures were unusual.

   D She liked to imagine her grandmother as a pioneer girl.

Qualitative Results: Reading [3]

Grade 5

- Revised only = 7
- Enhanced only = 3
- Revised and enhanced = 7
- Made no changes = 1

Grade 8

- Revised only = 17
- Enhanced only = 0
- Revised and enhanced = 0
- Made no changes = 1

Qualitative Results: Mathematics [1]

Math Strands & Overall Format

- Organized items by strand
- Ordered items to support engagement (e.g., easier items before more difficult items; simple applications of concepts before more complex applications)
- Organized strands to avoid having most difficult set of items at the end of the test
- Added more white space
  - Between responses, especially when visuals were involved
  - To break up dense text in multi-step problems
  - Separate question context from question to be answered

Qualitative Results: Mathematics [2]

- Revised Items by
  - Eliminating extraneous information, especially in the contexts for multi-step problems
  - Simplifying the vocabulary (e.g., use “2” for “two’)
  - Eliminating potential visual discrimination problems (e.g., should both “6” and “9” be used in a place value question?)
- Enhanced items by
  - Adding a “familiar” graphic (e.g., table of values) or symbol
  - Adding a hint or thought balloon (e.g., Hint: A *mean* is kind of average; don’t forget to simplify)
  - Adding a short/simplified example to get them thinking
  - Underlining, enlarging, or bolding key words or symbols

Sample Training Example for Mathematics

(reorganize & simplify text, underline key word)

BEFORE

Bill was discussing the total sales in his store this week. He said that on Monday there were twice as many sales as Thursday, less 13. If there were 135 sales on Thursday, how many sales did he have on Monday?

AFTER

Bill had 135 sales on Thursday. The sales on Monday were twice as many as Thursday, less 13.

How many sales did Bill have on Monday?

Sample Training Example for Mathematics

(change format; simplify reading load)

BEFORE

- Mary has six white shirts, four blue shirts, and five red shirts in her dresser. If she randomly chooses…

AFTER

Mary has

- 6 white shirts
- 4 blue shirts
- 5 red shirts

If she randomly chooses…

Qualitative Results: Mathematics [3]

Grade 5

- Revised only = 3
- Enhanced only = 1
- Revised and enhanced = 16
- Made no changes = 0

Grade 8

- Revised only = 3
- Enhanced only = 0
- Revised and enhanced = 17
- Made no changes = 0

Pilot Test

- Pilot test conducted in February 2008
  - Two test forms per grade
  - Items appeared in original state on one form and modified state on second form (forms counter-balanced)
  - Items identified as potential served as link items, appearing in original state on both forms

- Both regular and special education students participated
  - Grade 5: 3,741 students
  - Grade 8: 3,647 students

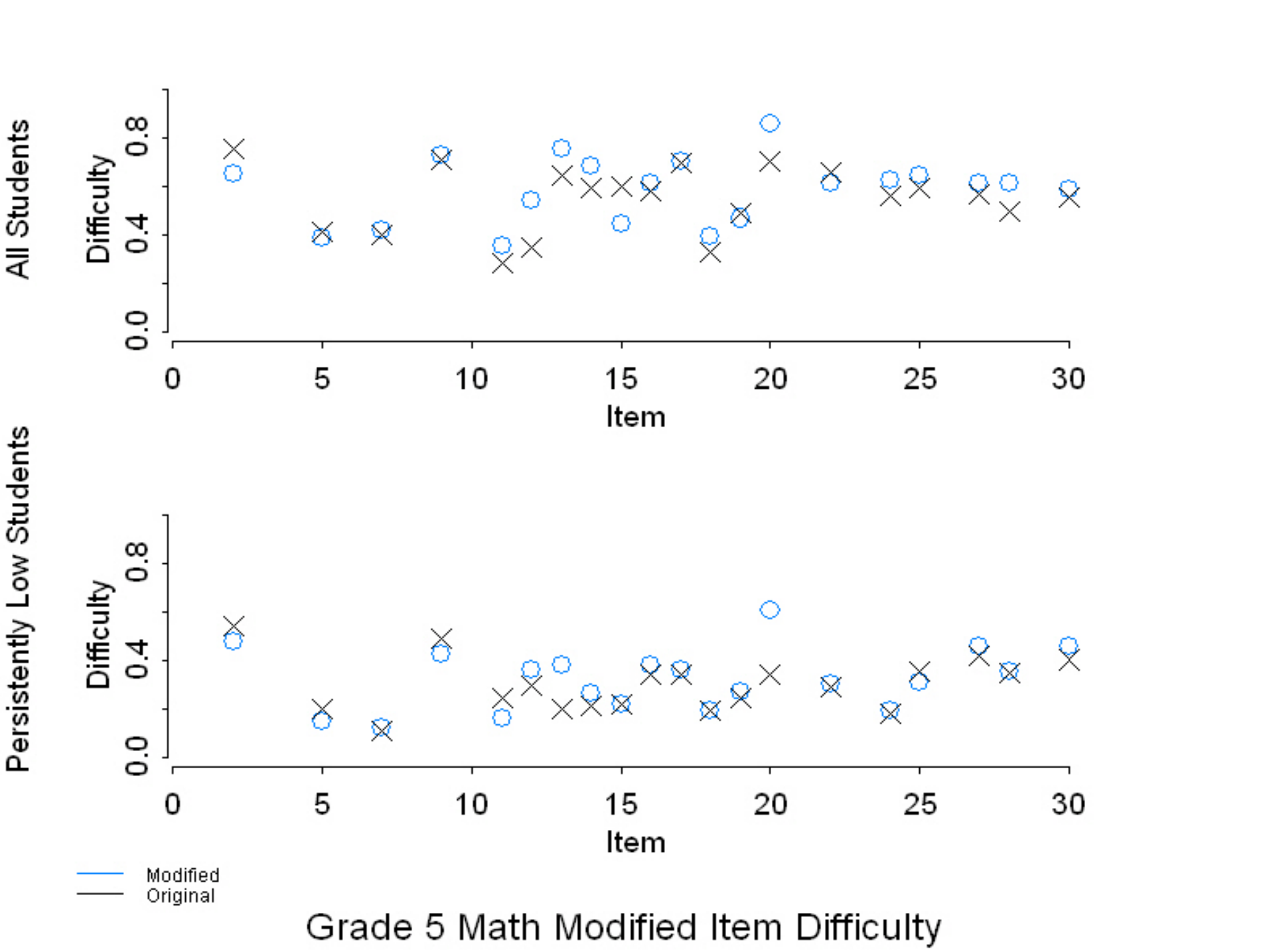
Summary of Original Identification

	Number of Items	Effective	Ineffective
Math: Grade 5	60	31%	29%
Math: Grade 8	60	23%	31%
Reading: Grade 5	40	35%	20%
Reading: Grade 8	40	15%	3%

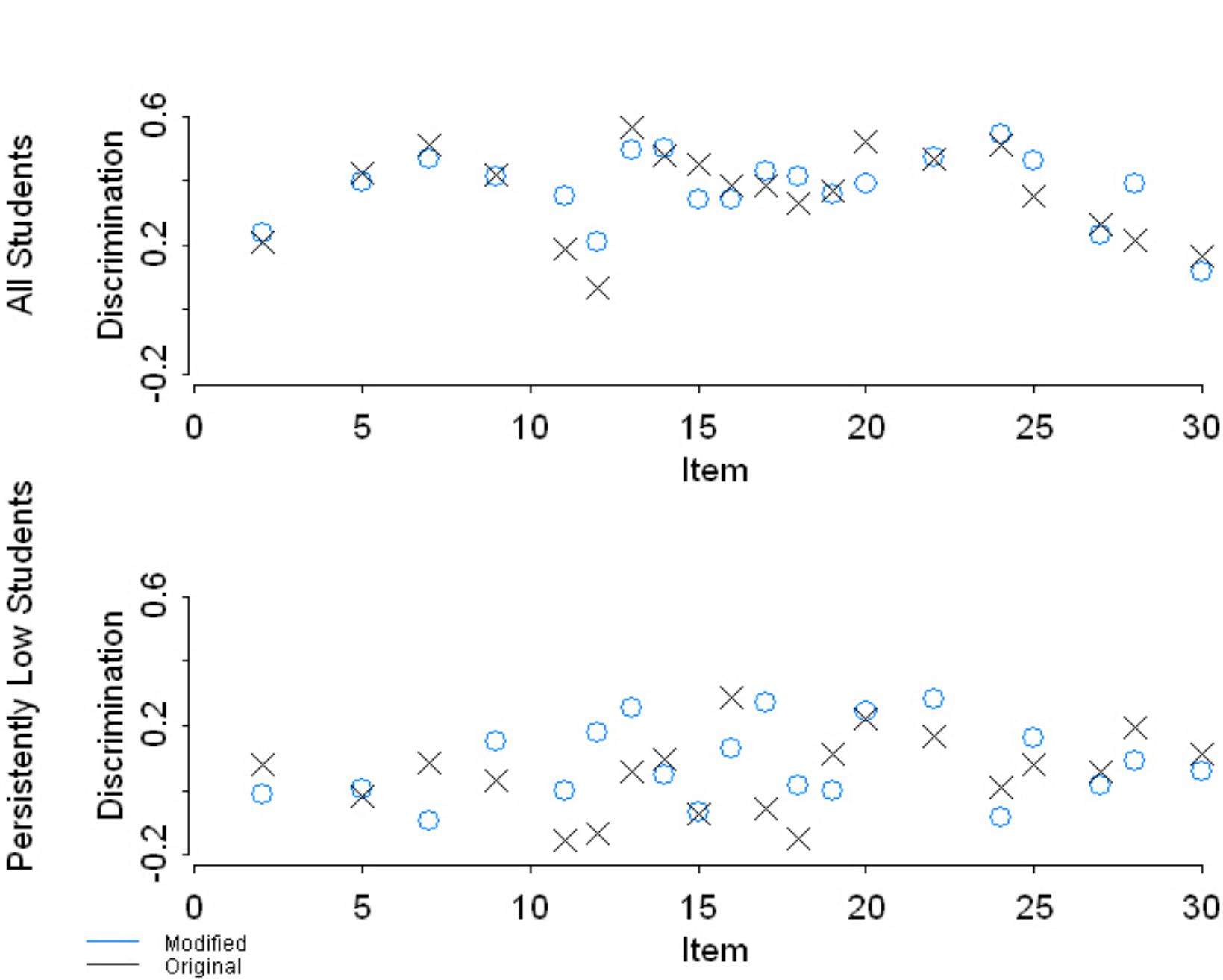
Summary of Pilot Identification

	Number of Items	Effective		Ineffective	
		Form 1	Form 2	Form 1	Form 2
Math: Grade 5	30	70%	53%	10%	23%
Math: Grade 8	30	63%	70%	33%	20%
Reading: Grade 5	25	76%	56%	8%	16%
Reading: Grade 8	25	68%	88%	16%	4%

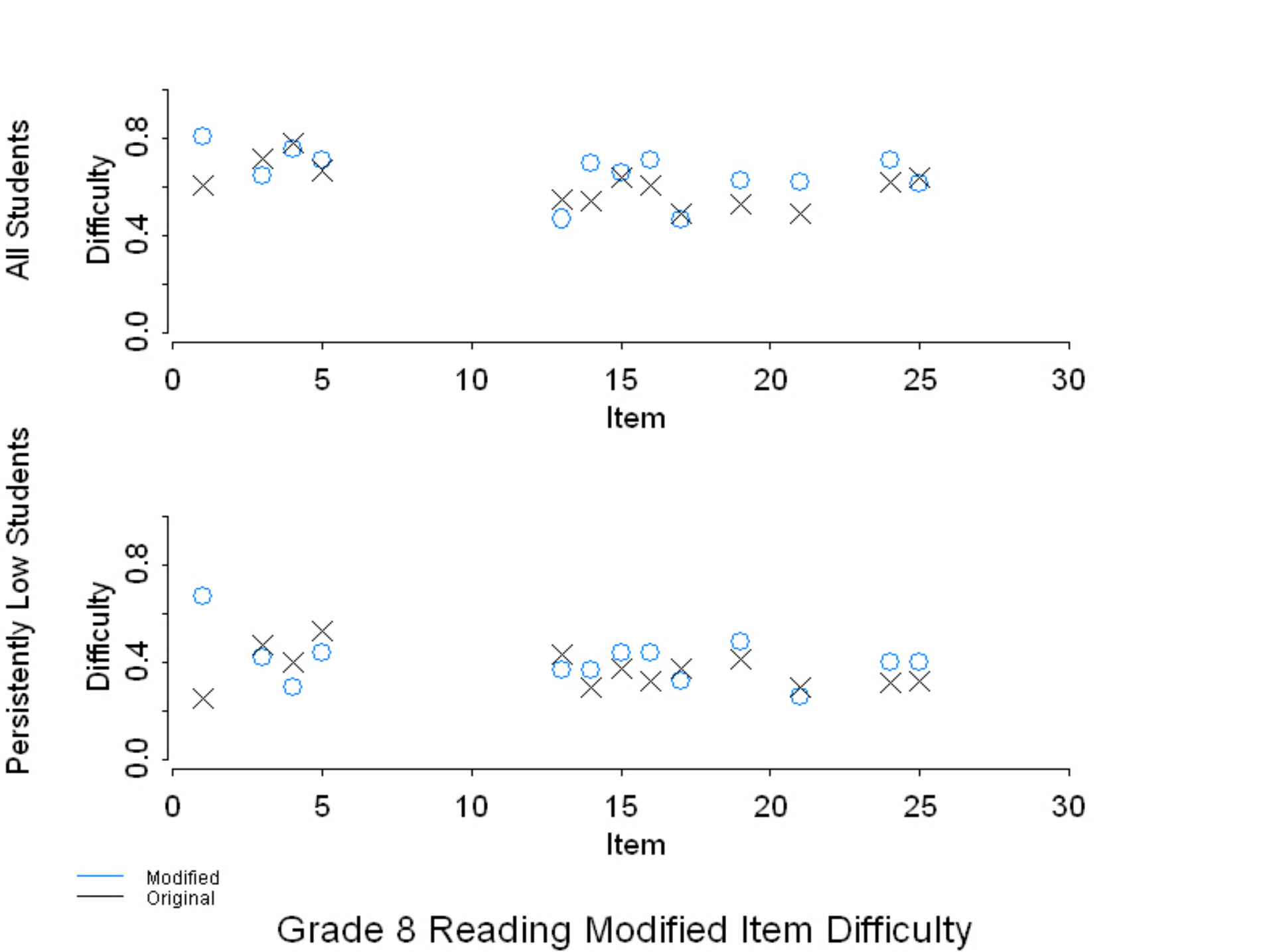
Item Position	Form 1			Form 2			Content Strand
	Modification	Diff	Disc	Modification	Diff	Disc	
1	Not Modified			Not Modified			G & M
2	Not Modified			Modified			G & M
3	Not Modified			Not Modified			G & M
4	Not Modified			Not Modified			G & M
5	Modified			Not Modified			N & O
6	Not Modified			Not Modified			N & O
7	Not Modified	X		Modified	X		N & O
8	Not Modified			Not Modified			N & O
9	Modified			Not Modified			N & O
10	Not Modified			Not Modified			N & O
11	Not Modified			Modified			SP
12	Modified			Not Modified		X	SP
13	Modified			Not Modified	X		PS
14	Not Modified			Modified	X	X	PS
15	Modified	X		Not Modified	X	X	PS
16	Not Modified			Modified			PS
17	Modified			Not Modified			PS
18	Modified			Not Modified			PS
19	Not Modified			Modified		X	PS
20	Not Modified			Modified			PS
21	Not Modified			Not Modified			PRA
22	Modified			Not Modified			PRA
23	Not Modified			Not Modified			PRA
24	Not Modified		X	Modified		X	C & E
25	Modified			Not Modified			C & E
26	Not Modified			Not Modified			C & E
27	Not Modified			Modified			C & E
28	Not Modified			Modified			C & E
29	Not Modified			Not Modified			C& E
30	Modified			Not Modified			C & E



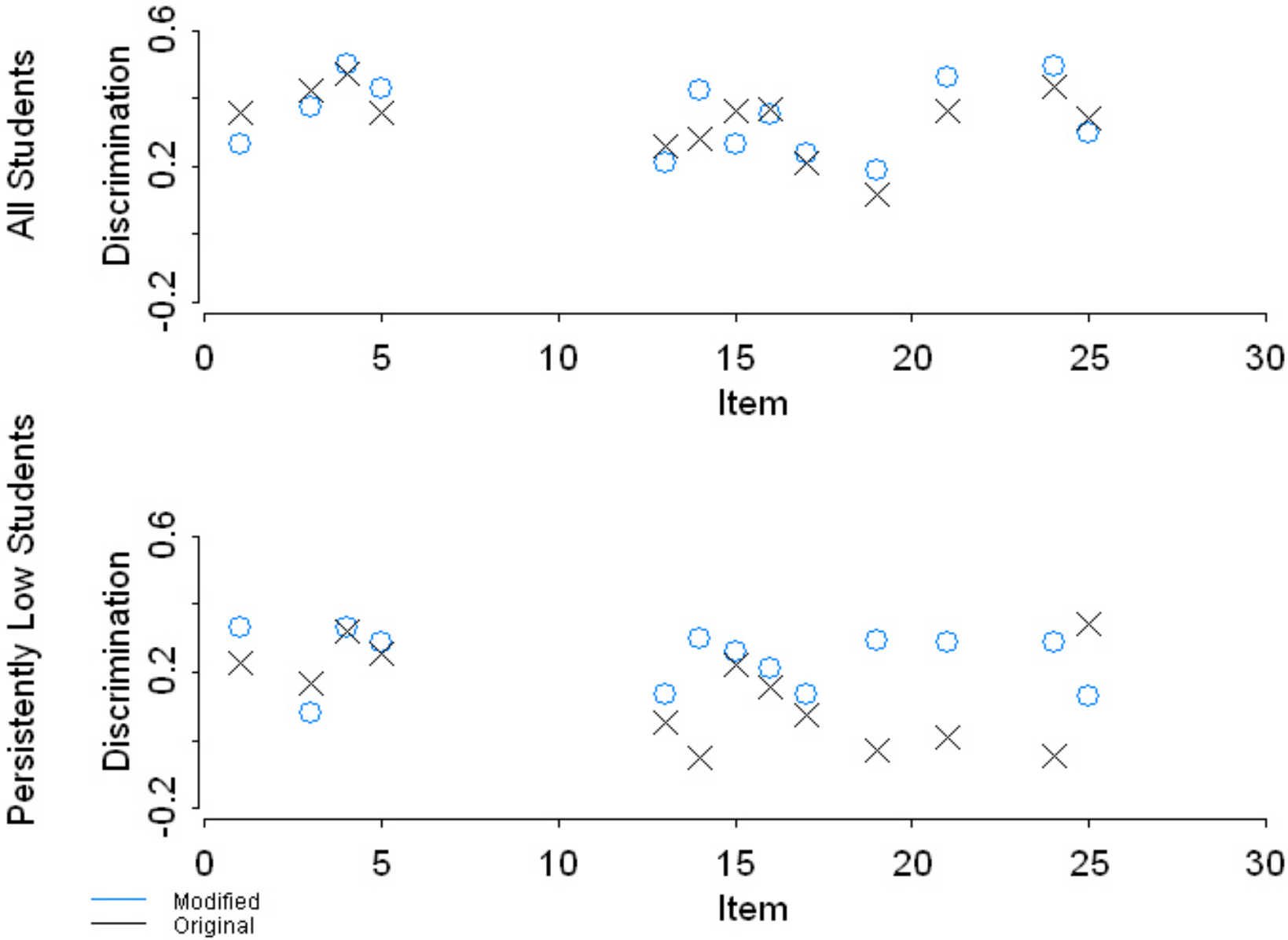
Grade 5 Math Modified Item Difficulty



Grade 5 Math Modified Item Discrimination



Grade 8 Reading Modified Item Difficulty



Grade 8 Reading Modified Item Discrimination

What have we learned?

- Statistical Techniques
  - Providing new insight into the effectiveness of item modifications
- Modifications
  - There does not appear to be a consistent effect
  - The type of modification may have an impact
  - Item by item effect
- Next Steps
  - IRT analyses
  - Investigate the relationship between the statistical results and the type of modification

Concluding Thoughts…

- Not all modifications will work for all students
  - Cognitive labs can help determine why some modifications might not work for some students
- Opportunity to learn must be attended to early and often
  - Classroom observations might shed additional/needed light on what is going on instructionally with identified students

Concluding Thoughts…

- Finding have implications for item development for general assessment as well
- This is hard work and there isn't going to be a simple solution
- Focus on the issue –
  - This endeavor is about better measurement of student achievement so that appropriate instructional decisions can be made