

Illinois Alternate Assessment (IAA)
2007-2008
Annual Technical Manual

August 14, 2008
Pearson

Table of Contents

<i>Introduction</i>	6
 <i>Chapter I Test Overview and Design</i>	
Historical Overview.....	8
Test Design Changes for 2007-2008.....	14
 <i>Chapter II Item Development</i>	
Item Writing Process	19
Item Specifications.....	19
Qualifications of Item Writers and Method of Recruitment.....	21
Training Practices/Activities (in consideration of both content and bias).....	22
Outcome of the Item Writing Process.....	22
 <i>Chapter III Test Materials and Teacher Training</i>	
Test Implementation Manual.....	25
Test Booklets.....	25
Answer Sheets.....	25
Online Test Platform.....	25
Teacher Training.....	25
 <i>Chapter IV Fall 2007 Item Field Test</i>	
Overall Pilot Test Plan.....	27
Initial Pilot Test.....	27
Sampling Plan.....	27
Forms Design.....	28
Analysis and Use of Pilot Test Data.....	29
Target Population.....	30
Field Test Form Assignment.....	31
Assignment and Placement of Events within the Seven Field Test Forms.....	31
Assignment of Pilot Forms.....	31

Analysis.....	32
Classical Item Analyses.....	33
Differential Item Functioning (DIF) Analyses.....	33
Reliability	34
Test Score Reliability	35
IRT Analysis.....	36
Item Calibration and Equating.....	36
Data Review.....	37
Bias Review.....	39

Chapter V Spring 2008 Operational Test Administration

Materials and training.....	41
Administration.....	41
Test forms.....	41
Embedded Field Test Items.....	42
Classical Item Analyses.....	42
Reliability.....	42
Calibration and Equating.....	43
IRT Analysis.....	45
Scaling.....	46

Chapter VI Standard Setting for the IAA Mathematics, Reading, Science, and Grade 6 Writing Tests

Introduction.....	48
Goal of the Standard Setting Panel	48
Overview of the Standard Setting Workshop.....	48
Body of Work Procedure.....	48
Item Mapping Standard Setting Process.....	49
Review of Impact Data.....	49
Vertical Articulation.....	49
Performance Level Definitions.....	50
IAA Standard Setting Workshop.....	50

Setting Three Performance Standards.....	50
Feedback to Guide Panelist’s Judgments.....	50
Standard Setting Panels.....	51
Recruiting Panel Members.....	51
Panel Composition.....	51
Workshop Procedures.....	52
Training.....	52
General overview of setting performance standards on student assessments.....	52
Orientation to the Body of Work standard setting procedure.....	52
Orientation to the item mapping standard setting procedure.....	53
Orientation to the IAA test materials and training process	54
Orientation to Illinois content standards and Performance Level Definitions.....	54
Standard Setting.....	54
Conducting the Body of Work Method.....	55
Reviewing the Ordered Item Booklet & Conducting the Item mapping Process.....	57
Presentation of Impact Data.....	59
Evaluation of Standard Setting Workshop.....	63
Vertical Articulation.....	63
Recommended Final Cut Scores and Impact Data.....	63
Conclusions.....	66

Chapter VIII Spring 2008 Item Field Test

Field Test Design.....	67
Forms Assignment	7
Analysis.....	67
Classical Item Analyses.....	68
Differential Item Functioning (DIF) Analyses.....	69
IRT Analysis.....	69
Data Review.....	69
Item Bank.....	69

Chapter IX Spring 2008 Item Field Test

Introduction.....	70
-------------------	----

Performance based measurement.....	70
IAA Reliability.....	72
IAA Validity.....	74
Methods.....	74
Results.....	75
Expert Score Inter-Rater Reliability.....	75
Correlations with Expert Scores.....	75
Discussion.....	76
References.....	77

Appendices

Appendix A. Sample IAA Task.....	A1
Appendix B. IAA Scoring Rubric.....	A3
Appendix C. IAA Paper Scoring Sheet.....	A5
Appendix D. Item Analysis – Spring 2008 Operational Forms.....	A7
Appendix E: 2008 IAA Standard Setting Item Maps.....	A66
Appendix F: 2008 IAA Standard Setting Evaluation Form	A99
Appendix G: 2008 IAA Standard Setting Evaluation Results.....	A104
Appendix H: Item Analysis – Spring 2008 Field Test.....	A117

Introduction

In 1997, the Illinois Standard Achievement Test (ISAT) was authorized by state law to measure how well students learned the knowledge and skills identified in the Illinois Learning Standards. The Illinois Alternate Assessment (IAA) was added to the assessment program in 2000 to meet the requirements of the Individuals with Disabilities Education Act of 1997 (IDEA) and the No Child Left Behind Act (NCLB) of 2001. These laws mandated that an alternate assessment be in place for those students with significant cognitive disabilities who are unable to take the standard form of the state assessment even with accommodations. Eligibility for participation in the IAA is determined by the student's Individualized Education Program (IEP) team. The original IAA was a portfolio based assessment. In 2006, Pearson was contracted by the Illinois State Board of Education (ISBE) to develop, administer and maintain a new IAA. The first subject area developed for this new assessment (i.e., Writing) was piloted in Fall 2006 and administered operationally in Spring 2007. Reading, Math, and Science subject areas for the IAA were developed and piloted in Fall 2007, and operationally administered in Spring 2008.

Purpose

The purpose of this 2008 IAA Technical manual is to provide objective information regarding technical aspects of the IAA tests. This volume is intended to be one source of information to Illinois K-12 educational stakeholders (including testing coordinators, educators, parents, and other interested citizens) about the development, implementation, scoring, and technical attributes of the IAA. Other sources of information regarding the IAA, provided in paper or online format, include the IAA Administration Manual, implementation material, and training materials.

The information provided here fulfills legal, professional and scientific guidelines (AERA, APA, NCME, 1999) for pilot test technical reports of large scale alternate educational assessments and is intended for use by qualified users within schools who use the IAA and interpret the results. Specifically, information was selected for inclusion in this report based on NCLB requirements and the following *Standards for Educational and Psychological Testing*:

- Standards 6.1 – 6.15 Supporting Documentation for Tests
- Standards 10.1—10.12 Testing Individuals with Disabilities
- Standards 13.1—13.19 Educational Testing and Assessment

This technical report provides accurate, complete, current and clear documentation of the IAA development methods, data analysis, and results as is appropriate for use by qualified users and technical experts. Chapters I - III provide an overview of the test design, test content, and test administration materials. Chapter IV details the Fall 2007 pilot test. Chapters V – IX describe the Spring 2008 operational IAA administration, analyses, standard setting, and validation study.

Information provided in this manual provides valuable information about the IAA regarding:

1. Content of the tests;
2. Test form design;

3. Identification of ineffective items;
4. Reliability of the tests;
5. Difficulty of the test questions;
6. Equating of test forms;
7. Test administration process and materials;
8. Appropriateness of the directions for the tests;
9. Detection of item bias;
10. Setting performance standard cut scores;
11. Scoring and reporting the results of the tests;
12. Student performance on the IAA across the state overall and by subgroup; and
13. Validity of the IAA.

Each of these facets in the IAA test development and use cycle is critical to validity of test scores and interpretation of results. This technical manual covers all of these topics for the 2007-2008 testing year.

Chapter I

Test Overview and Design

Test Overview

In this section we provide a historical overview of the best practices support, psychometric rationale, and evidence from peer reviews for other States that informed the approach implemented by Pearson and ISBE for IAA materials production and provision. The IAA is at the cutting edge of alternate assessment design. The assessment provides a number of improvements over portfolio and checklist methods through an events-based measurement approach. The events-based approach provides greater standardization of individualized alternate assessments than do other assessment methods. Improvements result in higher test reliability, broader validity evidence, greater alignment to content standards, a basis for more rigorous standard setting processes, and more rigorous bias detection methods. Additionally, feedback from the field on implementation of the IAA supports broad acceptance by special education experts, high levels of assessment material usability, and positive impacts of the assessment on instruction and classroom practices. This chapter is organized into two broad sections: The first provides background on the initial development of the test and the second reports changes and updates to the test design implemented during the 2007-2008 test cycle.

Historical Overview

Best practices in alternate assessment design

NCLB legislation has driven the development of alternate assessments as a part of large scale assessment in reading, math and science for the 1% population of students with disabilities who are not able to participate in a state's standardized assessments, even with accommodations. This is an important point regarding the 1% population and appropriate assessment design for these students: The population and the appropriate assessments are more individualized than are the standardized assessments for the general student population.

Whereas instruction and assessment are standardized by design for the general student population, students in the 1% population vary widely by disability type, and consequently: a) in mode of instruction, b) materials used in instruction, and c) materials used in assessment. However, for both populations, the most valid assessment is that linked closely to the mode and material used in student instruction. Appropriate instruction—given the population—is equally reliant on teacher execution for both populations and drives the validity of either type of assessment. Consequently, best practices in alternate assessment design mirror those for general assessments in that each provides an assessment experience that is most appropriate for the student, best aligned with student instruction, and provides the most valid and useful assessment results. For alternate assessments, this requires a design in which flexibility of assessment administration is not compromised for the sake of total standardization, because without flexibility the assessment will not be aligned with the student's disability and instruction; resulting in lowered assessment validity and usability of results.

Relevant psychometric research

The development, administration, and scoring plans for the IAA were based on the latest psychometric research on alternate assessments. A complete review of that research is beyond the scope of this brief; however, Schafer (2005) in his paper “Technical Documentation for Alternate Assessments” summarizes eight critical points regarding alternate assessment psychometrics:

1. Test every student on what he or she is supposed to be learning. That should be the primary focus of any alignment study (or process for ensuring alignment). In other words, remember the Fundamental Accountability Mission.
2. State each inference that is to be supported specifically in a psychometric evaluation of the validity of any assessment or assessment program. Elaborate each statement to address all relevant questions stakeholders may have, such as instructional implications, implications for certification of achievement, and institutional implications. Collect validity evidence for each inference separately in such a way to evaluate the assumptions that are necessary for the inference.
3. The most important inference to focus on for statewide assessments is that of assignments to achievement levels. This is also true for alternate assessments. Other inferences may also be important depending on the context. Unintended inferences may need study too, even if only to show that they are invalid.
4. For all assessments, and especially for alternate assessments, evaluate the reliability and validity for the student’s assessment results, for their referents, and for the process by which they are compared.
5. Assessments of reliability that focus on evidence across examinees (e.g., variance components, correlations) are probably **not** going to be useful in studying alternate assessments. Instead, document the consistency of the score and of its referent independently, as well as the process of making the comparison, and focus on the process as it occurs at the individual student level.
6. The student’s instructional domain must be consistent with criteria for alternate achievement standards and the student’s alternate assessment must be aligned with that instructional domain. Researchable aspects of these criteria are: (1) is the breadth of allowable individual content expectations sufficient to include the appropriate instructional domain for each student, (2) does the instructional domain provide access to all aspects of the regular curriculum, (3) does the student’s alternate assessment align with the student’s instructional domain, and (4) do the student’s performance expectations represent the highest possible achievement standards that are consistent with the student’s instructional domain.
7. On-demand assessments and their associated need for standardization may not be crucial for alternate assessments. Indeed, it may even be best to evaluate maximal rather than typical student performance. Decisions about whether to require standardized, on-demand data

collection or to generate data less formally might best be made at the individual student level (i.e., individualized).

8. Include within the criterion of utility, how well the assessment system provides explicit instructional focus for the teacher. Consider this recommendation as a companion to the Fundamental Accountability Mission.

NCLB requirements

In December 2003, the US Education Department released regulations allowing states to develop alternate achievement standards for students with the most significant cognitive disabilities. These standards had to have the same characteristics as grade-level achievement standards: they must be aligned with the State's academic content standards, they must describe at least three proficiency levels; reference the competencies associated with each achievement level; and include cut scores that differentiate among the levels. The regulations also stipulated that a recognized and validated procedure must be used to determine each achievement level.

States were not required to adopt alternate achievement standards. However, if they chose to do so, the standards, and the assessment used to measure students with the most significant cognitive disabilities against those standards would be subject to federal peer review. The August 2005 non-regulatory guidance on developing alternate achievement standards specified states could develop alternate assessments based on alternate achievement standards, but provided little guidance as to the format of these assessments, other than stipulating they must meet the same requirements as all other assessments under Title I, i.e. the same technical requirements as the regular assessment.

The non-regulatory guidance provides states significant latitude in designing the format of alternate assessments for alternate achievement standards. They specifically state that there is no typical format, reference the Title I regulations that require alignment to the state's content standards, and suggest that an alternate may reduce the breadth and or depth of those standards (Alternate Achievement Standards for Students with the Most Significant Cognitive Disabilities: Non-regulatory Guidance, August, 2005, p.16). Essentially, USED has indicated that it is most concerned with the technical adequacy of the alternate assessments and alignment with state content standards. Provided states follow best psychometric practices in developing their alternate assessments and document their processes, the format of any alternate assessment is secondary to the requirement to measure the content standards.

The most relevant NCLB requirements for the IAA were those that had been explicitly addressed to ISBE through the peer review letter. Points that were made regarding the IAA are provided below and have been addressed and documented in the work Pearson and ISBE have completed and/or planned under the current IAA contract:

4.0 - TECHNICAL QUALITY

5. Documentation of the technical adequacy of the Illinois Alternate Assessment (IAA):

The use of procedures for sensitivity and bias reviews and evidence of how results are used; and

Clear documentation of the standard-setting process.

5.0 – ALIGNMENT

5. Details of the alignment study planned for the IAA. This evidence should include the assurance that tasks used are appropriately aligned/linked to the academic performance indicators.

Feedback from other States

At the time the IAA was developed, sixteen states had received full approval or full approval with recommendations from the peer review process. Of those, six had task based alternate assessments. These states were Alaska, Colorado, Michigan, South Carolina, Utah, and West Virginia. All of these states had a specified testing window and pre-determined tasks, events, or items. The states differed in how many entry points, or levels of difficulty, they included in individual items, whether different items were selected for groups of students based on their functional levels, and whether proficiency was based on a sum score or scale score. However, none of these differences impacted whether the assessments were approved during peer review. The majority of these states allowed for items or events to be customized to the student's environment and classroom materials.

Several additional states that had approval pending status in the peer review process also had task based alternate assessments. None of these states failed to receive approval due to the format of the assessment. The primary reason alternate assessments were not fully approved in these states was a lack of evidence of the alignment to the state's content standards. A lack of technical documentation, standards setting processes, and achievement level descriptors were also common deficiencies.

Excerpts from August, 2005 non-regulatory guidance:

According to the December 9, 2003 regulation, and as determined by each child's IEP team, students with disabilities may, as appropriate, now be assessed through the following means:

The regular grade-level State assessment

The regular grade-level State assessment with accommodations, such as changes in presentation, response, setting, and timing. For more information about accommodations, see <http://education.umn.edu/NCEO/OnlinePubs/Policy16.htm>

Alternate assessments aligned with grade-level achievement standards

Alternate assessments based on alternate achievement standards.

The 2004 IDEA amendments reinforce that children with disabilities may be appropriately assessed through one of these four alternatives.

To qualify as an assessment under Title I, an alternate assessment must be aligned with the State's content standards, must yield results separately in both reading/language arts and mathematics, and must be designed and implemented in a manner that supports use of the results as an indicator of AYP. Alternate assessments can measure progress based on alternate achievement standards (see Section C) and can also measure proficiency based on grade-level achievement standards. Alternate assessments may be needed for students who have a broad variety of disabilities; consequently, a State may employ more than one alternate assessment.

When used as part of the State assessment program, alternate assessments must have an explicit structure, guidelines for which students may participate, clearly defined scoring criteria and procedures, and a report format that communicates student performance in terms of the academic achievement standards defined by the State. The requirements for high technical quality set forth in 34 C.F.R. §§200.2(b) and 200.3(a)(1), including validity, reliability, accessibility, objectivity, and consistency with nationally recognized professional and technical standards, apply to alternate assessments as well as to regular State assessments.³

What is the typical format for an alternate assessment?

There is no typical or single format for an alternate assessment. Some alternate assessments are built on portfolios of student work or activities that demonstrate knowledge through performance of specific tasks. An alternate assessment may include materials collected under a variety of circumstances, including (1) teacher observation of the student; (2) samples of student work produced during regular classroom instruction that demonstrate mastery of specific instructional strategies; and (3) standardized performance tasks produced in an "on-demand" setting, such as completion of an assigned task on test day. These are not requirements. They are only examples of different types of alternate assessments. States have considerable flexibility in designing the most appropriate format for alternate assessments.

An alternate assessment based on alternate achievement standards may cover a narrower range of content (e.g., cover fewer objectives under each content standard) and reflect a different set of expectations in the areas of reading/language arts, mathematics, and science than do regular assessments or alternate assessments based on grade-level achievement standards. The questions on an alternate assessment might be simpler than those on a regular assessment or the expectations for how well students know particular content standards may be less complex but still challenging for students with the most significant cognitive disabilities. If a State chooses to use such assessments, it must establish alternate achievement standards through a documented standards-setting process; the assessments based on alternate achievement standards must yield separate results for reading/language arts, mathematics, and (beginning in the 2007-08 school year) science. Proficient and advanced scores in reading/language arts and mathematics from an alternate assessment based on alternate achievement standards may be used in AYP decisions in the same manner as any other scores, subject to the 1.0 percent cap at the LEA and State levels.

IAA Test Program Development

ISBE contracted Pearson, and their subcontractor partners, the Inclusive Large Scale Standards and Assessment (ILSSA) group, and Beck Evaluation and Testing Associates, Inc. (BETA) in 2006 to develop the new IAA in grades three through eight and 11 for Reading and Mathematics; in grades four, seven, and 11 for Science; and in grades three, five, six, eight, and 11 for Writing. The Pearson team, working with ISBE and the Assessment Committee for Students with Disabilities (ACSD), developed an events-based assessment that includes performance tasks to best measure achievement through links to the Illinois Learning Standards. A sample IAA task is provided in Appendix A. An events-based assessment provides more objective measurement than does a portfolio based alternate assessment, and requires less teacher and student time to administer. Several factors were taken into consideration during planning and development of the IAA program including:

- The IAA will reflect the breadth and depth of content of the tested content areas and grade level.
- The IAA will promote access to the general curriculum.
- The IAA will reflect and promote high expectation and achievement levels.
- The IAA will allow access to students with the most significant cognitive impairments, including those with sensory impairments.
- The IAA will be free from racial, gender, ethnicity, socioeconomic, geographical region, and cultural bias.
- The IAA will not increase the teachers' burden to assess and is non-obtrusive to the instructional process.
- The IAA will meet federally mandated requirements.

In addition to being based on instructional activities in the general curriculum, the assessment task development utilized the theory and elements of Universal Design for Learning. Specifically, multiple means of expression and representation were addressed. In addition, a BETA alternate assessment design specialist recommended instructional/assessment strategies that could be used effectively with the content and design.

Scoring of the events-based assessment for the pilot and operational administrations is conducted in the classroom by a qualified and trained teacher, using a scoring rubric as a part of the alternate assessment program to categorize student performance into four specific performance levels. Teachers observe the students with significant cognitive disabilities one-on-one during the administration of the six assessment events. The observation of each assessment task is scored according to the rubric. The rubric for the IAA was developed in collaboration with the ISBE, the ACSD, and educators. These rubrics were used during the scoring of the pilot administration of the IAA to determine if the definitions and criteria were adequately delineated to facilitate reliable scoring. The rubric was evaluated for appropriateness and technical quality and has been modified in collaboration with ISBE and the development team using pilot data and input from the field. The rubric is provided in Appendix B.

Test Content

In the spring of 2006, a team of Illinois educators created the new Illinois Alternate Assessment (IAA) Frameworks. These Frameworks are located on the ISBE website at www.isbe.net/assessment/iaa.htm. The purpose of the frameworks is to prioritize the skills and knowledge from the Illinois Learning Standards in order to develop a new Illinois Alternate Assessment for students with the most significant cognitive disabilities. ILSSA's responsibilities included facilitating the development of the IAA Frameworks and providing statewide staff development on how to access grade-level curriculum. Pearson has continued to refine the IAA frameworks.

The focus of the content standards alignment work was directed to the alignment of the IAA Frameworks to the general education grade level academic standards. When Pearson began work on development of the IAA, ISBE had in place the Illinois Learning Standards in a number of content areas to guide curriculum and assessment for all students. Success for students on all Illinois tests, including the IAA, required teaching students to the Illinois Learning Standards. Sometimes it was difficult for educators to determine how a student taking the IAA can be taught and assessed on the Illinois Learning Standards. In order to build content validity into the IAA, Pearson and ILSSA aligned the IAA Frameworks to the Illinois Learning Standards. One goal of developing the IAA Frameworks in this way was to provide a content valid assessment framework; an equally important goal was to provide an instructional resource for special education teachers aligned to the Illinois Learning Standards.

The first step of this process was facilitated by ILSSA and completed by Illinois educators who determined the critical function of each assessment objective which would be used to write the IAA Frameworks. The critical function was what educators expect ALL kids to know or do in order to meet an assessment objective and was written in common language so that someone who is unfamiliar with the content area can understand the purpose of the objective. Determining the critical function of a standard was one way to assist educators to determine this. ILSSA trained a group of educators to assist in development of the IAA Frameworks by starting with the intent of the standard; providing examples of how a variety of students can access the standard, related curricula and materials; and then defining the critical function based on this work. The educators were reminded that students taking the IAA will receive instruction on grade level content standards (may be at a lower complexity level) within the context of grade level curriculum ensuring that the intent of the grade level content standard remains intact through the alignment process. The critical functions used for the alignment are provided in the IAA Frameworks document.

Based on the alternate assessment research and best practices covered earlier in this section, and the foundational test content work described here, Pearson commenced development of the IAA tests. In the following sections the item development, forms building, item field testing, operational administration, scoring and validation processes and results are described for the new IAA.

Test Design Changes for 2007-2008

In response to feedback from teachers administering the IAA and concerns over the level of item standardization across students, Pearson, at ISBE's direction, modified many of the IAA items to provide greater standardization and decrease ambiguity and confusion in item administration protocol. In order to strike a balance between standardization and flexibility in testing the 1% population IAA items were written to two broad formats: general and specific. In the general item format, more flexibility was given to the teacher to design the task to best suit each individual student. These items were written to require that students perform a specific skill, but provided an example of how the item might be administered rather than specifically dictating the task. An example of a general format item might be, "Provide the student with three single-digit whole numbers and ask the student to identify one of the numbers. Example: Provide the student with the whole numbers: 5, 8, and 9. Ask the student, 'Which number is 8?'. In the specific item format, the task expected of a student was explicitly documented and the teacher was directed to administer it exactly as specified while taking into account each student's individual needs (e.g., administer the item in accordance with the student's mode of communication). An example of a specific item might be, "Provide the student with the whole numbers: 5, 8, and 9. Ask the student, 'Which number is 8?'"

Teachers administering the IAA were provided with training on how to differentiate between, and administer, general versus specific items. Despite this training, feedback from the field after the Fall 2007 Pilot study overwhelmingly showed that teachers administering the test had difficulty determining if an item was intended to be specific or general. The result of this was that the majority of teachers chose to administer general items as specified in the example, thereby negating the added flexibility offered by this format. Additionally, ISBE felt that the general items, particularly in the case of the Reading tests, allowed too much range in the materials that students might be presented for a single item.

In response to these concerns, ISBE requested that Pearson examine differences between these two types of items using data from the Fall 2007 Pilot Study. Pearson examined differences between the two item types using both classical test theory (i.e., item mean, item-total correlation, item score distribution) and IRT based item statistics (i.e., average ability of students earning each item score point and model in-fit). Results of the analyses showed that the different item categories showed no substantive difference in terms of these statistics. The results of the classical analyses are presented in Table 1.1 and the results of the IRT analyses are shown in Table 1.2.

Table 1.1. Classical Test Theory Comparison of General vs. Specific Items

			Item Mean	Item Total Correlation	Percent 1	Percent 2	Percent 3	Percent 4
Math	Grade 3	General	2.90	0.76	14	21	18	46
		Specific	2.88	0.73	17	21	16	47
	Grade 4	General	2.99	0.77	14	17	15	53
		Specific	3.00	0.74	14	15	17	54
	Grade 5	General	2.89	0.78	15	20	18	48
		Specific	2.97	0.77	13	19	18	50
	Grade 6	General	3.02	0.79	14	17	18	51
		Specific	3.06	0.77	11	19	20	50
Reading	Grade 7	General	2.84	0.78	13	24	19	44
		Specific	2.72	0.78	15	28	20	38
	Grade 8	General	2.84	0.79	16	24	16	44
		Specific	2.80	0.76	19	23	19	39
	Grade 11	General	2.84	0.76	11	20	24	46
		Specific	2.86	0.74	12	19	21	47
	Grade 3	General	3.04	0.78	12	17	18	54
		Specific	3.04	0.77	9	20	18	53
Science	Grade 4	General	3.09	0.79	12	13	19	55
		Specific	3.14	0.79	10	15	16	59
	Grade 5	General	3.02	0.79	12	16	21	52
		Specific	3.09	0.78	10	17	20	54
	Grade 6	General	3.11	0.77	12	16	21	52
		Specific	3.09	0.75	10	17	20	54
	Grade 7	General	3.13	0.80	10	15	16	59
		Specific	3.15	0.80	9	15	18	59
Science	Grade 8	General	3.20	0.80	10	15	18	57
		Specific	3.11	0.80	10	18	18	54
	Grade 11	General	3.30	0.80	6	9	16	69
		Specific	3.39	0.80	7	9	13	71
Science	Grade 4	General	2.81	0.79	18	20	17	45
		Specific	2.88	0.77	14	21	19	46
	Grade 7	General	3.07	0.77	12	17	14	57
		Specific	2.90	0.77	13	22	19	46
	Grade 11	General	3.12	0.78	9	11	19	61
		Specific	3.12	0.79	9	13	18	60

Table 1.2. IRT Comparison of General vs. Specific Items

			Category Average 1	Category Average 2	Category Average 3	Category Average 4	Item Fit
Math	Grade 3	General	-1.82	-0.04	0.71	1.80	0.99
		Specific	-1.69	0.02	0.83	1.75	1.00
	Grade 4	General	-1.74	0.08	0.87	1.95	0.97
		Specific	-1.68	0.17	0.96	1.89	1.04
	Grade 5	General	-2.05	-0.02	0.87	2.00	1.00
		Specific	-2.04	-0.03	0.86	1.95	0.97
	Grade 6	General	-1.97	0.00	1.02	2.19	0.99
		Specific	-2.09	0.01	0.88	2.15	1.00
	Grade 7	General	-2.27	-0.20	0.91	1.91	1.02
		Specific	-2.19	-0.05	1.03	2.00	0.97
	Grade 8	General	-2.39	-0.16	0.91	2.01	0.98
		Specific	-2.27	-0.12	0.91	1.86	1.08
	Grade 11	General	-2.15	0.07	0.95	1.78	0.98
		Specific	-2.18	-0.06	0.82	1.60	1.05
Reading	Grade 3	General	-1.92	0.03	0.99	2.11	0.99
		Specific	-1.95	0.11	0.99	2.04	1.02
	Grade 4	General	-1.83	0.05	1.10	2.20	1.01
		Specific	-1.92	0.04	1.07	2.21	1.00
	Grade 5	General	-2.20	0.15	1.18	2.28	0.99
		Specific	-2.30	0.00	1.03	2.20	1.01
	Grade 6	General	-1.91	0.01	0.97	2.04	0.98
		Specific	-1.91	0.13	0.90	2.01	1.02
	Grade 7	General	-2.16	0.06	1.12	2.50	1.01
		Specific	-2.38	-0.07	1.13	2.46	0.98
	Grade 8	General	-2.36	0.02	1.14	2.44	0.97
		Specific	-2.23	0.13	1.27	2.47	1.01
	Grade 11	General	-2.28	0.22	1.26	2.72	1.01
		Specific	-2.39	0.10	1.10	2.60	0.98
Science	Grade 4	General	-2.02	-0.07	0.93	1.93	0.96
		Specific	-2.06	-0.03	0.96	1.92	1.01
	Grade 7	General	-2.00	-0.23	0.64	1.85	1.02
		Specific	-1.86	0.10	0.99	2.01	0.99
	Grade 11	General	-1.97	0.05	0.91	1.96	0.97
		Specific	-1.90	-0.10	0.75	1.94	1.02

Given that the results of this study showed virtually no difference in the performance of these two item types ISBE requested that Pearson modify all general format items to align with the specific item format. This was typically a minor change that consisted of changing the example portion of the general format item to a prompt like those of the specific items. Both ISBE and Pearson psychometricians agreed, based on the results of the analyses reported previously, that

this modification would not significantly alter the Fall 2007 Pilot results such that they would be unusable for data and bias review. A more cautious approach, however, was taken with respect to the item statistics that would be used for operational scoring. The IAA was originally intended to be a pre-equated test with the item statistics derived from the pilot studies (Fall 2006 and Fall 2007) being used for scoring. Because the modification of items was conducted after these pilot administrations ISBE and Pearson psychometricians deemed it necessary to move the IAA to a post-equating model for all tests (Mathematics, Reading, Science, & Writing - the post-equating process is described in Chapter 5 of this Technical Manual). The base scale for Mathematics, Reading, and Science had yet to be established at this point. In light of this it was decided that item statistics from the Fall 2007 Pilot of these tests would not be submitted to the item bank or reported in this Technical Manual. Instead, only item statistics for items administered operationally or in field-test positions from the Spring 2008 and future administrations would be included in the item bank.

In order to strike a balance between test length and content coverage the IAA was originally designed with four required items and one teacher selected item. The four required items measured the objectives identified as priorities by Pearson content staff, and the teacher selected item allowed a teacher administering the test to select one additional item that was most appropriate for each student, based on that student's IEP, from a set of items representing the remaining objectives. Based on the results of the Spring 2007 operational Writing test, it was clear that the majority of teachers were selecting the first option among the choices. This showed that the teacher selected item was not functioning as expected. ISBE then requested that Pearson increase the length of the Mathematics, Reading, and Science tests and remove the teacher selected item option from all tests. ISBE chose to lengthen the Mathematics test to ten required items, the Reading test to nine required items, and the Science test to six required items. In order to better facilitate item replenishment for these lengthened tests two items were field tested on each form for these subjects. The Writing test was not lengthened, although the teacher selected option was removed leaving the test with five required items and one field test item.

Chapter II

Item Development

Item Writing Process

Based on the content frameworks described in the previous section, Pearson/ILSSA based IAA assessment tasks and passage development on a rigorous multi-step process that included the following components. This process is consistent with that used for other Illinois programs.

1. **Information gathering** – reviewed ISBE’s documentation, attended planning meetings, synthesized item/task and test specification, and determined plans for releasing tasks.
2. **Project-specific document creation** – developed project development plans and content- and state-specific task writer training materials.
3. **Task writer recruitment and training** – recruited and trained potential writers on industry best practices and IAA-specific styles and task requirements. ISBE, reviewed training, preparation, and presentation materials and participated in face-to-face, web-based, and/or conference call training.
4. **Task Development** – procured tasks; review and editing of tasks performed by content and alternate assessment specialists to address content accuracy, alignment to curriculum and/or test specifications, principles of Universal Design, grade and cognitive level appropriateness, level of symbolic communication, scorability with the rubric, and language usage; copy was edited for sentence structure, grammar, spelling and punctuation; created art; evaluated tasks for potential bias/sensitivity concerns; and reviewed source and accuracy
5. **Independent Review** – reviewed by ILSSA content specialists for overall task quality and alignment to ISBE's *Guidelines for Test Development* and the test specifications.
6. **Initial customer review** – reviewed by, and received feedback from, ISBE staff on a sampling of approximately 20 tasks per subject early in the development cycle to check for a common understanding of ISBE expectations for quality and for content and cognitive mapping.
7. **Committee reviews** – review of passages and tasks by Illinois stakeholders for content and bias/sensitivity with Pearson/ILSSA staff. Qualitative reviews of passages and tasks took place before field test and quantitative reviews (data review) takes place after the pilot administration of the tasks.

Item Specifications

The following is a general description of the Illinois student population being assessed by the IAA. This description was used as context for item development purposes only. These students have, or function as if they have, **significant** cognitive disabilities. Students in this population most likely:

- Have both physical and mental disabilities, and

- Use an alternate form of communication

These students exist along a disability continuum—some students may have one of the more severe forms of Autism, some may have Downs Syndrome and others may be multiply cognitively and physically impaired in ways that severely limit their ability to function in the classroom.

Based on this understanding of the population to be tested, the IAA events, activities, and stimuli were written in accordance with Universal Design principles, which emphasize the maximization of readability and comprehensibility (from Synthesis Report 44)¹:

1. Simple, clear, commonly-used words should be used, and any unnecessary words should be eliminated.
2. When technical terms must be used, they should be clearly defined.
3. Compound complex sentences should be broken down into several short sentences, stating the most important ideas first.
4. Only one idea, fact, or process should be introduced at a time; then develop the ideas logically.
5. All noun-pronoun relationships should be made clear.
6. When time and setting are important to the sentence, place them at the beginning of the sentence.
7. When presenting instructions, sequence steps in the exact order of the occurrence.
8. If processes are being described, they should be simply illustrated, labeled, and placed close to the text they support.

By applying writing and editing guidelines that promote clarity in language, style, and format, the IAA assessments maximize accessibility so students may better show what they know and are able to do. Following best practices in item writing for alternate assessments and the Universal Design philosophy, writers and editors were directed to adhere to strategies such as those outlined in the Table 2.01.

¹ Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal design applied to large scale assessments* (Synthesis Report 44). Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes. Retrieved August 19, 2003, from the World Wide Web: <http://education.umn.edu/NCEO/OnlinePubs/Synthesis44.html>.

Table 2.01. Plain Language Editing Strategies (from Synthesis Report 44)

Strategy	Description
Reduce excessive length.	Reduce wordiness and remove irrelevant material. Where possible, replace compound and complex sentences with simple ones.
Eliminate unusual or low frequency words and replace with common words.	For example, replace “utilize” with “use.”
Avoid ambiguous words.	For example, “crane” could be a bird or a piece of heavy machinery.
Avoid irregularly spelled words.	For example, “trough” and “feign.”
Avoid proper names.	Replace proper names with simple, common names such as first names.
Avoid inconsistent naming and graphic conventions.	Avoid multiple names for the same concept. Be consistent in the use of typeface.
Avoid unclear signals about how to direct attention.	Well-designed headings and graphic arrangement can convey information about the relative importance of information and order in which it should be considered. For example, phrases such as “in the table below,…” can be helpful.
Mark all questions.	When asking more than one question, be sure that each is specifically marked with a bullet, letter, number, or other obvious graphic signal.

Qualifications of Item Writers and Method of Recruitment

The majority of item writers were Illinois Special Educators, whose names were provided to us by ISBE. A few of the teachers were from other states, had successfully written items for their respective state’s severely cognitively impaired student population, and had experience writing items in the past. All item writers had experience teaching the population being measured. In Table 2.02 is provided the number of item writers who worked on the IAA tasks by subject.

Table 2.02. Number of IAA item writers per test subject

Subject	Number of item writers
Mathematics	9
Reading	8
Science	6

Training Practices/Activities (in consideration of both content and bias)

Item writer training was held on October 17, November 2, and November 6, 2006 for Science, Mathematics, and Reading, respectively. Prior to the item writer training, training materials were shipped to each writer. The materials included the following: a cover letter explaining what was included in the materials packet, the IAA frameworks (including priorities and associated worksheets) for the grade for which the teacher would write items, item template, sample items, a security oath/contract, some information regarding the development of the frameworks (priorities and their associated worksheets), sample items, and a blueprint that explained the number of items to develop per priority assessment objective. During the item writer training, materials were reviewed in detail and the writers were trained on how they were to use the materials when developing the items. A general description of the population being assessed by the IAA and the manner in which the students would be assessed were also provided. Item writers were trained in using the IAA frameworks as part of item writer training.

Outcomes of the Item Writing Process

Tasks created during the 2006 item writing efforts were reviewed by grade level panels of reviewers February 26-29, 2007. These reviewers were trained in the following areas:

- How to use the Frameworks in conjunction with the Items under review,
- Student Expectation—Item Match
- Appropriateness of Item for the Severely Cognitively Impaired
- Adequacy of Preparation
- Freedom from Bias

Each reviewer was provided with a copy of the IAA Frameworks to use when reviewing the items. They were asked to consider how the item matched to the assessment objective for which it was written. Panel members were led through a page by page review of each item and allowed ample time for the reviewers to discuss each item. Panel members had the opportunity to recommend that ISBE accept an item as presented, revise the item on the spot, or recommend as a group to “DNU” (Do Not Use) the item. After each item review, item editors made the changes suggested by the committee members and proofed the items against the notes taken during item review. The results of this item review are as follows.

Math Results:

- Of the Grade 3 math items, 0 items were approved without modification, 64 items were approved after modification, and 2 items were deleted.
- Of the Grade 4 math items, 0 items were approved without modification, 64 items were approved after modification, and 2 items were deleted.
- Of the Grade 5 math items, 0 items were approved without modification, 62 items were approved after modification, and 3 items were deleted.
- Of the Grade 6 math items, 0 items were approved without modification, 62 items were approved after modification, and 4 items were deleted.

- Of the Grade 7 math items, 0 items were approved without modification, 60 items were approved after modification, and 7 items were deleted.
- Of the Grade 8 math items, 0 items were approved without modification, 61 items were approved after modification, and 5 items were deleted.
- Of the Grade 11 math items, 0 items were approved without modification, 61 items were approved after modification, and 4 items were deleted.

Reading Results:

- Of the Grade 3 reading items, 0 items were approved without modification, 51 items were approved after modification, and 0 items were deleted.
- Of the Grade 4 reading items, 0 items were approved without modification, 54 items were approved after modification, and 0 items were deleted.
- Of the Grade 5 reading items, 0 items were approved without modification, 59 items were approved after modification, and 5 items were deleted.
- Of the Grade 6 reading items, 0 items were approved without modification, 61 items were approved after modification, and 4 items were deleted.
- Of the Grade 7 reading items, 0 items were approved without modification, 58 items were approved after modification, and 1 item were deleted.
- Of the Grade 8 reading items, 0 items were approved without modification, 59 items were approved after modification, and 4 items were deleted.
- Of the Grade 11 reading items, 0 items were approved without modification, 45 items were approved after modification, and 10 items were deleted.

Science Results:

- Of the Grade 4 science items, 0 items were approved without modification, 58 items were approved after modification, and 1 item was deleted.
- Of the Grade 7 science items, 0 items were approved without modification, 56 items were approved after modification, and 3 items were deleted.
- Of the Grade 11 science items, 0 items were approved without modification, 52 items were approved after modification, and 0 items were deleted.

At ISBE's direction two additional tasks were undertaken following the item review meetings. Additional artwork was created for some Mathematics and Science tasks in order to clarify for the teacher what was needed for that particular task. Specific teacher instructions (and no examples) were provided for 60% of the items (per grade and subject area) and 40% of the items were provided with general teacher instruction and specific examples. This process led to the distinction between general and specific items discussed in Chapter 1 of this Technical Manual.

By the end of this process, Pearson had developed and taken through item reviews sufficient numbers of IAA tasks to build an item pool ready for field testing that would sufficiently cover the content frameworks. Many of these items were included in the Fall 2007 field test.

Because exposure of the IAA tasks is a concern, as it is for any large scale assessments with stakes for students and/or schools, tasks that are used once operationally are not used again, if at all possible. In order to support this design, Pearson replenishes the item bank with newly written items each year and field tests these items in an embedded design. When the IAA for Mathematics, Reading, and Science became operational in Spring 2008, 14 field test items were field tested in each grade with two items embedded in each of seven versions of the operational form.

Chapter III

Test Materials and Teacher Training

Given that the IAA design includes teacher administration and scoring of the test for a population of students who each have unique set of needs, test material design and teacher training are very important to the validity of the test. All test materials have gone through extensive development and review processes by Pearson and ISBE. Furthermore, materials have gone through a pilot testing and a validation study. Training was developed by Pearson in collaboration with ISBE to be delivered at regional settings across Illinois, and to be provided via a web-based solution.

Test Implementation Manual

The IAA test implementation manual was developed by Pearson for ISBE using input from best practices and the field. Within the test implementation manual the teacher can find all information necessary to prepare for, administer, and provide scores back to Pearson for the IAA. Additionally, links to teacher training material for the IAA are also included in the manual to be used as a refresher course. The IAA implementation manual is available online at www.isbe.net/assessment/iaa.htm.

Test Booklets

IAA test booklets are very much like test booklets for general assessment tests, except they contain fewer items. Each test booklet contains a set of operational items and subset of field test items. Items are scored using the four point rubric described in an earlier section of this report and provided in Appendix B.

Answer Sheets

The IAA answer sheets have been developed by Pearson and ISBE to be user friendly, efficient means of data capture. The answer sheet can be located at the back of each test booklet. Teachers record the student's scores on the answer sheet during test administration and then transfer the scores to the online platform at a later time. The paper answer sheet is provided in Appendix C of this report.

Online Test Platform

Pearson *SchoolSuccess* Group provides an online platform for teachers to use in IAA score submission. Training for the online platform is provided by Pearson to teachers and test coordinators statewide. The online platform speeds data collection and minimizes student identification errors.

Teacher training

Training Objectives

- Increase participants' familiarity with IAA calendar of events and timeline expectations
- Improve participants' understanding of the Illinois Learning Standards and IAA Frameworks

- Promote scoring reliability and validity through practice exercises using the newly devised IAA rubric
- Present video clips of students engaged in the IAA to explore educators' rationale for score assignment and test preparation efforts
- Detail best practices for test administration including assessment procedures, emphasis on students' primary mode of communication, materials modification, and creating optimal testing environments
- Offer guidelines for materials modification including the receipt, verification and return of secure test materials
- Demonstrate capabilities of online scoring tool

Training Logistics

- Throughout August and September of 2007, Pearson, in partnership with ISBE and the Assessment Committee for Special Education (ACSD), conducted 16 onsite trainings in 8 locations statewide in preparation for the Fall 2007 Pilot and Spring 2008 operational assessment
- Each session was attended by approximately 100 Illinois IAA Coordinators and educators
- In addition to these onsite trainings, Pearson lead 9 statewide webinars with an emphasis on the online scoring process
- Pearson structured its conference calls to specific audiences including Chicago Public Schools (CPS), Special Education Cooperatives (co-op), and private facilities

Training Facilitators

- Each onsite session was introduced by one-two ISBE team members including an overview of the IAA test design, ownership of policy decisions, and availability of CPDUs
- Each onsite session was facilitated by a Pearson IAA team representing both the Program and Content Support Services teams

Training Materials

- All materials in support of the IAA Regional Training program and Spring 2008 test administration were developed by Pearson in consultation with and approval from ISBE
- Materials were accessible to educators via the ISBE IAA website at www.isbe.net/assessment/iaa.htm and/or distributed to Illinois educators in conjunction with IAA's Spring 2008 packaging and distribution requirements
- Regional Training materials included an 85-slide PowerPoint presentation, IAA rubric, practice scoring activity to enable evaluation of student video clips, sample answer document to acquaint participants with required data fields that were used in the spring 2008 operational
- Test administration resources included the IAA Frameworks, the 30-page Test Implementation Manual, Online User Guides for Teachers, Coordinators and Secondary Scorers, and test books

Chapter IV

Fall 2007 Item Field Test

The Fall 2007 pilot test of Mathematics, Reading, and Science items was designed to mirror the Fall 2006 pilot test conducted for Writing. As discussed in Chapter 2 of this Technical Manual, changes to items from these content areas following this pilot administration and prior to operational administration in Spring 2008 led to a decision not to retain the item statistics derived from this field test for operational scoring and item banking. Nonetheless, the results of the pilot test were deemed appropriate for use in data and bias review. As such, this chapter will provide an overview of the field test process, analysis, and subsequent data and bias review. In keeping with the decision not to retain item statistics from this administration, however, specific results will not be reported. Chapter 5 of this Technical Manual provides detailed item statistics for items from this field test chosen for operational administration in Spring 2008 and Chapter 8 provides detailed statistics for embedded field test statistics from that administration.

Overall Pilot Test Plan

The overall pilot test design used by Pearson was an initial stand-alone, census pilot test in which only newly developed pilot test events were administered (i.e., no scored events were included in the forms) to all eligible students. This initial pilot test is followed by embedded pilot testing of events within operational forms each year for replenishment of events over time.

Initial Pilot Test

The initial pilot test provided a test of the events and the system to see how well various aspects of psychometric and measurement properties perform and to provide evidence regarding valid score use, reliability and fairness/appropriateness. Logistically, events were assigned to teachers and students statewide who used an online system to record the student scores regarding student attainment of the skills measured by the events. Administration of the pilot test, like the operational test, was conducted by educators trained on the IAA process by Pearson and was conducted with students on an individual basis. Only certified teachers used the IAA scoring rubric, yielding scores from 1 to 4 for each task. Unlike the operational test, teachers administered a set of six events to each student in the initial pilot test. The initial pilot test of IAA Mathematics and Reading events for grades 3- 8 and 11 and Science events for grade 4, 7, and 11 was conducted in Fall 2007.

Sampling Plan

The initial pilot test was a census design; it included all eligible students in the State. This design had several advantages over a purposeful pilot test sampling plan, including:

- Due to the small population size, we would not be able to collect enough alternate assessment students from a sample pilot to perform statistical calibration.
- A census pilot gave all the schools an opportunity to familiarize themselves with the new IAA content areas before they would have to administer them operationally.
- It avoided confusion regarding why some districts were selected and some were not.

- It avoided the perception that schools with large numbers of alternate assessment students were “penalized” through required participation in the pilot.

According to ISBE data for portfolios processed in Spring 2006 and IAA Writing tests submitted in Spring 2007, the expected number of students taking the IAA was approximately 1,500 per grade level. Because the pilot test occurred in the Fall of the school year, but the operational administrations are scheduled for Spring, the decision was made to conduct a recalibration of the item bank. This was based on a post-equating study rather than a pilot test of students one grade above. The items were targeted to control for student maturation and learning effects that would skew the initial item parameters from the pilot test.

Forms Design

Based on the statistical requirements of the Rasch-Partial Credit Model for item parameter estimation, a minimum of $n = 200$ student responses was needed for each task in the pilot test. Given the expected numbers of students per grade available to participate in the pilot test (i.e., ~1,500) Pearson developed and administered seven pilot test forms per grade spiraled across administration settings (e.g., schools). Advantages to spiraling include: a) even distribution of pilot test forms across the student population, and b) equivalence of groups across items for the IRT analysis (i.e., randomly equivalent groups). Advantages to spiraling across administration settings include less burden on teachers administering the items (i.e., the same events would be given to all students within a grade by a given teacher) and ease of forms distribution (i.e., each administration setting receives a single form of pilot test items to administer to all students in a grade level).

The number of items that had been written, reviewed and were available for the pilot test varied by grade and subject. At each grade, the items covered the essential goals in the IAA Frameworks in a fairly even manner. Given the number of students available for the pilot test, and limitations on pilot test form length, not all items could be included in the initial pilot test. Total numbers of items that could be included was dependent on the final pilot test form design.

Initial operational form design for the IAA evolved based on input from ISBE and their stakeholders, alternate assessment content specialists, and Pearson technical input. Pilot forms basically mirrored the 2007 Writing operational forms and the intended 2008 operational form design at the time.

1. Each operational IAA form will include five scored items and one embedded field test item. Of the five operational items, four are required and one is chosen by the teacher from a list of optional items. Seven forms will be administered per subject per grade ($n = 200$ per form) with one unique field test item in each form. All students were administered the same four required items within a given administration year.
2. Four required priority objectives will be identified per subject per grade and the operational form will represent these with four required items every year. This option gives comparability across years within grade and a basis for Adequate Yearly Progress (AYP).
3. Required priorities within subject will differ across grades to the extent that is appropriate in order to adequately cover a subject area domain across all grades (i.e., 3 – 8, 11).

4. Required items will be retired after one operational administration and replaced with psychometrically sound field tested items from the same priorities.
5. All priorities not identified as one of the four required will be represented by optional items for the operational administration, given availability of existing psychometrically sound items for these priorities.
6. Optional items will be recycled based on the least operational exposures per item. Those items that have been chosen the fewest number of times for administration will be recycled first.
7. The four required items and the embedded field test item will be rotated through the first five operational form positions across the seven forms per grade, with the item representing the optional priorities always in the final position on the form.

ISBE approved the use of a six- item form design for the initial pilot test, with five operational items and one embedded pilot test item intended for subsequent operational forms. This form design allowed 35 items to be included in the initial pilot test. Each form included items sampled from different primary goal areas as identified by Pearson content specialists and ISBE.

Each of the seven pilot test forms per grade was linked to two of the other six through common linking items. This design provided a mechanism for concurrent calibration of different items across forms. Common linking items were expected to have approximately 400 responses each and items unique to each form were expected to have approximately 200 responses each.

Analysis and Use of Pilot Test Data

Using the data collected from the pilot-tested items, Pearson performed a statistical calibration of the incomplete data matrix that resulted from the pilot study. Although not all students were evaluated on all items during the pilot, all pilot tested items had some responses and all students engaged in six items. As such, the dataset used for statistical calibration was similar to the one presented in the figure below.

	Event Number							Student
Student	1.3.01	1.3.02	1.3.03	•	•	•	2.3.10	Ability
S ₁								1.25
S ₂								-0.23
S ₃								4.57
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
•	•	•	•	•	•	•	•	•
S _j								1.58
Event								
Difficulty	-1.32	3.45	2.87				-0.09	

This calibration used the Rasch Partial-Credit Item Response Theory model. This measurement model, after calibration, placed all items and all students on the same underlying scale. As such,

future use of the items had a direct link back to this scale and direct comparability of student performance.

All analyses resulting from the pilot testing and the concurrent calibration of the incomplete data matrix were reviewed, documented and provided in a database for the ISBE to inspect. Some of the results from the analyses included the following Classical Test Theory statistics:

- **Item Means.** The mean raw item score will be computed for each polytomous item. Blank responses, legitimate zero (0) item score points and non-scorable responses were distinguished, as specified by the ISBE.
- **Item Score Point Frequency Distributions.** These data provide information about the effectiveness of the scoring rubric and how well the overall item functions across the years.
- **Corrected Item-to-Total Score Correlations.** This statistic helps evaluate how well an item discriminates between high-performing and low-performing examinees.

The following IRT analyses were also completed for all items.

- **Item Fit Estimates.** The extent to which the Rasch-Partial Credit model conformed to the data was estimated item by item. This diagnostic information was helpful in the selection of equating items eligible for psychometric equating.
 - **INFIT MNSQ** is the average of the INFIT mean-squares associated with the responses in each category. The expected values for all categories are 1.0.
- **Category step values.** These values are on the theta scale and represent the difficulty associated with scoring in one category versus the next lower category. This parameter indicates how difficult it is to observe a category, not how difficult it is to perform it.
- **Average measure.** The "average measure" for a category is the average ability of the people who scored in that category. $\text{Average measure} = \text{sum}(B_n - D_i) / \text{count of observations in category}$. This is an empirical value. It is not a Rasch-model parameter.
- **Test Information Functions.** An analysis was completed using IRT item difficulty estimates and ability thresholds to construct a test information function. This function indicates where, along the ability continuum, the discrimination and information about examinees is maximized, ideally around any performance level cut points. The utility of test information functions should be capitalized on during the process of test construction, and before test administration.

Items not meeting expectations were eliminated and/or revised based on data reviews with ISBE experts and stakeholders. Surviving items are the pool of items used in the future IAA tests.

Target Population

Students with significant cognitive disabilities in grades three through eight and 11 who are unable to take the ISAT even with accommodations or modifications take the IAA. Participation in and eligibility for the IAA is determined by the student's Individualized Educational Program (IEP) team. This group of students is referred to as the 1% population in NCLB guidelines.

Field Test Form Assignment

The Pilot test plan approved by ISBE delineated all critical points in the Pilot Test design. These points include:

- The Pilot was an initial stand-alone, census pilot test in which only newly developed pilot test items were administered (i.e., no scored items will be included in the forms) to all eligible students (i.e., the expected number of students taking the IAA is approximately 1,500 per grade level),
- The initial pilot test of IAA items was conducted in Fall 2007 with on-grade, eligible students,
- Seven forms of pilot test items were administered per grade with a target of $n = 200$ student responses for each task,
- Each of the seven pilot test forms per grade included six items and were linked to two of the other six forms through common linking items, and
- Each form included items sampled from across primary goal areas as identified by Pearson content specialists and ISBE.

Assignment and Placement of Items within the Seven Field Test Forms

Designs for: a) assigning items to field test forms, and b) assigning the forms to students, must work hand-in-hand in order to meet psychometric goals of calibrating all IAA items within a content area onto the same underlying scale. The field test forms were designed with linking items to control for differences in nonequivalent groups taking each form; therefore, strictly random assignment at the student level is not required. However, the quality of data from the linking items is critical to the design of the field test. Data quality is driven by true score variance and item reliability within the field test process. There must be adequate true score variance in the linking items (i.e., neither too easy nor too hard) to provide useable data across all students. Reliability of the linking items must be high enough to provide data that can be used for linking purposes. Additionally, these qualities of the items need to be equivalent across the forms they link. Pearson CSS and PS teams have worked together to review item content, cognitive complexity, and estimated difficulty when choosing the most appropriate linking items and assigning these items to appropriate positions within forms. The products of these efforts are reflected in the item maps for the Pilot test.

Assignment of Pilot Forms

Advantages to assigning pilot test forms in any effective spiraling design included: a) even distribution of pilot test forms across the student population, and b) equivalence of groups across events for the IRT analysis (i.e., randomly equivalent groups). Advantages to spiraling across administration settings included less burden on teachers administering the items (i.e., the same items would be given to all students within a grade by a given teacher) and ease of forms distribution (i.e., each administration setting receives a single form of pilot test items to administer to all students in a grade level).

The design of the pilot test forms allowed for assigning pilot forms to nonequivalent groups and mitigated the need for spiraling forms at the student level; however, it was necessary to maintain

equivalent sample size and subgroup representation across forms in order to ensure that satisfactory psychometric analyses of the pilot items could be conducted. Pearson determined that spiraling forms at the home school level would meet the requirement for adequate representation, as long as no single home school provided a large number of students. A preliminary analysis of n-counts for IAA portfolios processed by home school indicated that several home schools had within grade student counts higher than $n = 20$, which is approximately 10% of the targeted sample per form. In order to avert the potential impact these large sub-samples would have on the representation of students across forms, Pearson devised a spiraling plan that controlled for this effect. Based on guidelines provided by ISBE for teacher student ratios, Pearson determined that a 1:20 ratio would be appropriate for packaging and assigning pilot test forms. Specifically, any home school with 20 or fewer students within a grade was assigned a single form and were shipped a single packet with one pilot test form, one administration manual, and 20 answer sheets. Any home school with more than 20 students will be assigned an additional, different pilot test form for each 20 students. This provides a solution where forms were primarily spiraled across home schools, except in cases where the number of students within a grade in the home school is greater than 20, and then more than one pilot form will be spiraled within the home school. This forms assignment design resulted in equivalent numbers of students assigned to each pilot test form and equivalent representation of minority students and differences in school size across forms.

Analysis

Following the processing of student data, student demographic and item response data were transmitted to Pearson's psychometric services division. Pearson psychometric staff had primary responsibility for analyzing IAA field test data to ensure accuracy and validity of scoring. Most of the psychometric work was carried out using SAS Version 9.1 and WINSTEPS Version 3.6, commercially available statistical analysis software. Traditional item analysis and data file QC analyses were conducted with SAS programs. Item response theory (IRT) analyses were conducted with the WINSTEPS program (Linacre, 2006). WINSTEPS allows for estimation of IRT item parameters for dichotomously or polytomous scored items. It has been thoroughly tested and is currently utilized by several high-stakes testing programs administered by Pearson.

All technical support and analyses were carried out in accordance with both the Standards for Educational and Psychological Testing, issued jointly by the American Educational Research Association, the American Psychological Association, the National Council on Measurement in Education, and the Pearson Quality Assurance Program. Pearson staff verified the IAA data and analysis process at several steps in the procedure. This included verification of the SAS and WINSTEPS programs prior to use on actual pilot data through review by a second member of the psychometric services staff, and by using simulated data sets. Additionally, the output from the traditional and IRT item analysis programs were verified for out of range values and for consistent results across programs. Finally, the IRT calibrations were rerun independently by a second Pearson staff member.

Pearson conducted extensive statistical analyses on all pilot items. These analyses showed which items were at an appropriate difficulty level for the testing population and screened for differential item difficulty for subgroups in the student population. The analysis of the test data can be broken down into several components: 1) classical item analyses; 2) differential item

functioning (DIF) analyses; 3) reliability analyses; and 4) calibration of items. In the following sections, the analysis procedures for each component are described in detail.

Classical Item Analyses

Classical item analyses involve computing, for every task on each form, a set of statistics based on classical test theory. Analyses were reported both overall and by various ethnic and gender groups as sample size permitted.

The classical item statistics that were calculated for the pilot test items within each form include:

- Number of students tested for each item, overall and by subgroup.
Item Means. The mean raw item score was computed for each polytomous item and is analogous to the p-value for dichotomously scored items. This is a measure of the difficulty of an item, in classical test theory, and is indicated by the average raw score for an item across all students from the rubric ratings. For polytomously scored items, this statistic indicates the average rating earned on the item.
- Item Score Point Frequency Distributions. These data provide information about the effectiveness of the scoring rubric. Rubric point use by item was negatively skewed in every case.
- Corrected Item-to-Total Score Correlations. This statistic helps evaluate how well an item discriminates between high-performing and low-performing examinees. It is sometimes referred to as a discrimination index because it is an indicator of the degree to which students who do well on this content area also do well on this item. Items with negative or extremely low correlations can indicate serious problems with the item itself or can indicate that students have not been taught the content.

Differential Item Functioning (DIF) Analyses

One of the goals of the IAA test development is to assemble a set of items that provides a measure of a student's ability that is as fair and accurate as possible for all subgroups within the population. Differential item functioning (DIF) analysis refers to procedures that assess whether items are differentially difficult for different groups of examinees. DIF procedures typically control for overall between-group differences on a criterion, usually total test scores. Between-group performance on each item is then compared within sets of examinees having the same total test scores. If the item is differentially more difficult for an identifiable subgroup when conditioned on ability, the item may be measuring something different from the intended construct. However, it is important to recognize that DIF-flagged items might be related to actual differences in relevant knowledge or skills or statistical Type 1 error. As a result, DIF statistics are used to identify potential sources of item bias. Subsequent review by content experts and bias/sensitivity committees are required to determine the source and meaning of performance differences. In the IAA DIF analyses, DIF statistics were estimated for all major subgroups with sufficient sample size: Black, Hispanic, and Female. Items with statistically significant differences in performance were flagged so that items could be carefully examined for possible biased or unfair content that was undetected in earlier fairness and bias content review meetings held prior to form construction.

We used two statistical indices to identify DIF in the IAA pilot. First, we used the Mantel-Haenszel statistic provided from the Winsteps program output. A second type of DIF index, the

standardized mean difference across groups was computed by calculating the effect size difference (i.e., Cohen's d) for each subgroup compared to the majority group Cohen (1988) defined d as the difference between the means, $M_1 - M_2$, divided by standard deviation, s , of either group. In practice, the pooled standard deviation, s_{pooled} , is commonly used (Rosnow and Rosenthal, 1996). The pooled standard deviation is found as the root mean square of the two standard deviations (Cohen, 1988, p. 44). That is, the pooled standard deviation is the square root of the average of the squared standard deviations. When the two standard deviations are similar the root mean square will be not differ much from the simple average of the two variances.

$$d = M_1 - M_2 / \sigma_{\text{pooled}}$$

$$\sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$$

Cohen defined effect sizes as small, $d = .2$, medium, $d = .5$, and large, $d = .8$. These guidelines were used in interpreting the standardized mean differences for each item across subgroups and provided input for the task flagging procedure employed here. Items were flagged for DIF between a focal group (Black, Hispanic, Female) and the referent group (White, Male) according to five rules:

1. If the significance test provided for Mantel-Haenszel statistic by the Winsteps program for an item was not significant at the $p < .05$ level, *and* if the Cohen's d effect size was less than medium, the item received a flag value of "0"
2. If the significance test provided for Mantel-Haenszel statistic by the Winsteps program for an item was significant at the $p < .05$ level, *or* if the Cohen's d effect size was medium, but not large, the item received a flag value of "1"
3. If the significance test provided for Mantel-Haenszel statistic by the Winsteps program for an item was significant at the $p < .05$ level, *and* if the Cohen's d effect size was medium, but not large, the item received a flag value of "2"
4. If the Cohen's d effect size was large, *and* the Mantel-Haenszel statistic was *not* significant at the $p < .05$ level the item received a flag value of "2"
5. If the Cohen's d effect size was large, *and* the Mantel-Haenszel statistic was significant at the $p < .05$ level the item received a flag value of "3"

These DIF flag levels were defined by Pearson as follows:

0 = No Indication of DIF

1 = Slight Indication of DIF

2 = Possible Indication of DIF

3 = DIF Indicated

Reliability

The reliability of a test provides an estimate of the extent to which an assessment will yield the same results when administered in different times, locations, or populations, when the two administrations do not differ in relevant variables. Reliability coefficients are usually forms of correlation coefficients and must be interpreted within the context and design of the assessment

and of the reliability study. The forms of reliability below measure different dimensions of reliability and thus any or all might be used in assessing the reliability of IAA.

Test Score Reliability

Reliability focuses on the extent to which differences in test scores reflect true differences in the knowledge, ability, or skills being tested rather than fluctuations due to chance or factors other than those being tested. The variance in the distributions of test scores—essentially, the differences among individuals—is partly due to real differences in the knowledge, skills, or ability being tested (true variance) and partly due to random errors in the measurement process (error variance). The number used to describe reliability is an estimate of the proportion of the total variance that is true score variance.

$$\text{Reliability} = \sigma^2_{(\text{true score})} / \sigma^2_{(\text{total observed})}$$

Several different ways of estimating this proportion exist. The estimates of reliability reported in this report are internal-consistency measures, which are derived from analysis of the consistency of the performance of individuals on items within a test (internal consistency reliability).

Therefore, they apply only to the test form being analyzed. They do not take into account form-to-form variation due to equating limitations or lack of parallelism, nor are they responsive to day-to-day variation due, for example, to state of health or testing environment. Reliability coefficients may range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely individuals would be to obtain very similar scores upon repeated testing occasions with parallel forms. When the goal is to estimate the precision of a set of test scores from a single administration, a measure of internal consistency is frequently used to estimate reliability.

$$\alpha = (k/(k-1)) * [1 - \sum(s_i^2)/s_{\text{sum}}^2]$$

This is the formula for the most common index of reliability, namely, Cronbach's coefficient *alpha* (α). In this formula, the s_i^2 's denote the variances for the k individual items; s_{sum}^2 denotes the variance for the sum of all items. If there is no true score but only error in the items (which is esoteric and unique, and, therefore, uncorrelated across subjects), then the variance of the sum will be the same as the sum of variances of the individual items. Therefore, coefficient *alpha* will be equal to zero. If all items are perfectly reliable and measure the same thing (true score), then coefficient *alpha* is equal to 1. (Specifically, $1 - \sum(s_i^2)/s_{\text{sum}}^2$ will become equal to $(k-1)/k$; if we multiply this by $k/(k-1)$ we obtain 1.)

Several factors can affect reliability coefficients: 1) test length, 2) speededness, and 3) variance of true-scores. Test length is one factor that will affect both true-score variance and observed-score variance. In general, scores based on longer tests are more reliable due to the fact that as tests increase in length, true score and observed score variance increase faster than error score variance increases. Moreover, a longer test provides for broader sampling of the content domain, and thus more accurately reflects a student's performance on the domain as a whole.

As noted above, the magnitude of a reliability coefficient also depends on variation among students on both their true-scores and error scores, because reliability is a property of the scores

on a test for a particular group of examinees. Simply stated, as variance of true-scores decreases, reliability also decreases.

Coefficient alpha reliability was calculated for each IAA pilot test form.

IRT Analysis

Pearson estimated IRT parameters for all IAA pilot items to establish the underlying theta scale for each category level of each item. These parameter estimates serve to calibrate all students and test items onto the same underlying scale. The Rasch Partial Credit Model (RPCM) was selected because of its flexibility in accommodating a smaller n-count and for its ability to handle multiple-response category data. It also maintains a one-to-one relationship between derived scores (i.e., scale scores) and the underlying raw score scale. It is the underlying Rasch scale that facilitates equating of multiple test forms and allows for comparison of student performance across years. Additionally, the underlying Rasch scale facilitates the critical maintenance of equivalent performance standards across the years. The RPCM is defined via the following mathematical measurement model where, for a given item involving m score categories, the probability of person n achieving score x on prompt i is given by:

$$P_{xni} = \frac{\exp \sum_{j=0}^x (B_n - D_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (B_n - D_{ij})},$$

where, $x = 0, 1, 2, \dots, m$, and,

$$\sum_{j=0}^0 (B_n - D_{ij}) \equiv 0$$

The RPCM provides the probability of a student scoring x on the m_i step of question/prompt i as a function of the student's proficiency level B_n (i.e., sometimes referred to as 'ability') and the step difficulties (D_{ij}) of the m steps in prompt i . The data resulting from the calibration of the operational test using the RPCM (i.e., the scaled Rasch item and step difficulties) will be used to generate derived or converted scale scores when items are selected for use during subsequent operational assessment. Once performance standards are established, future use of any subset of these calibrated items will generate comparable (i.e., equated) results back to the first year, when the standards were established, thereby making the reporting of student scores comparable across the years.

Item Calibration and Equating

The purpose of item calibration and equating is to create a common scale for expressing the difficulty estimates of all the items across versions within a test. The scale commonly has a mean score of 0 and a standard deviation of 1. It should be noted that this scale is often referred to as the "theta" metric and is not used for reporting purposes because the values typically range from -3 to +3. Therefore, following calibration and equating, the scale is typically transformed to a reporting scale which can be meaningfully interpreted by students, teachers, and other stakeholders. All IRT analyses were conducted using the commercially available program Winsteps 3.6 (Linacre, 2006).

The following IRT analyses were completed for all items.

- Item Fit Estimates. The extent to which the Rasch-Partial Credit model conformed to the data was estimated item by item. This diagnostic information was helpful in the selection of equating items eligible for psychometric equating.
 - INFIT MNSQ is the average of the INFIT mean-squares associated with the responses in each category. The expected values for all categories are 1.0.
- Category step values. These values are on the theta scale and represent the difficulty associated with scoring in one category versus the next lower category. This parameter indicates how difficult it is to observe a category, not how difficult it is to perform it.
- Average measure. The "average measure" for a category is the average ability of the people who earn a given score point. $\text{Average measure} = \text{sum}(B_n - D_i) / \text{count of observations in category}$. This is an empirical value. It is not a Rasch-model parameter.

Data Review

Background

Data review represents a critical step in the test development cycle. At review meetings, panels of stakeholders in the IAA process, as well as ISBE staff and their representatives, had the opportunity to review actual student performance on the newly developed and field tested assessment items. The data review focused on the content validity, curricular alignment, and statistical functioning of field tested items prior to item selection for operational test forms. The field test results used in the data review provided evidence that the items were designed to yield valid results and were accessible for use by the widest possible range of students including students with significant types of disabilities or with limited English proficiency. The review of student performance should provide evidence regarding the fulfillment of requirement 200.2(b)(2) of NCLB. The purpose of the review meeting was to ensure that only psychometrically sound, fair, and aligned items are used in the construction of IAA test forms. Given that the reviewers attending the meetings provided their input on decisions regarding which items to keep for future operational form use, a clear explanation about the content of the items, the field test process, the scoring process, and the resulting field test data was critical to the success of these meetings and to the defensibility of the program.

Data review meetings were a collaborative effort between ISBE and Pearson. ISBE recruited panelists for the data review meetings from educators and other stakeholders in the IAA process; Pearson psychometricians and content specialists facilitated the meetings and trained committee members on how to interpret and review the field test data. Pearson verified the data review committees were beneficial and productive through informal debriefing with the panelists at the end of each session, as well as by collection of any post hoc comments provided by the panelists to ISBE or Pearson.

Meeting materials included the alternate content frameworks for the IAA, resources used for administering and scoring items, and review booklets of the items themselves with accompanying statistics generated from the field test data. Pearson content specialist provided background and training to the reviewers regarding the IAA content specifications, and answered

questions regarding item content during the meetings. Specific directions regarding the use of statistical information and the review booklets was provided by a Pearson psychometrician, who also confirmed that committee members had an appropriate level of understanding of the data and answered questions from the committee as they arose. Review of the data included presentation of item difficulty overall and by subgroups of students, item to total student score correlations, and indications of item DIF. Tasks failing to meet the requirements of sound technical data were carefully considered for rejection by the review panels and not used in future assessments; thereby enhancing the reliability and improving the validity of the items left in the system that will be used. While the panel used the data as a tool to inform their judgments, the panel (and not the data alone) made the final assessment as to the appropriateness or fairness of the assessment items.

Process

Operationally, the entire data review procedure was divided into three phases:

1. Pre-meeting preparation,
2. item content and statistics review meeting, and
3. post-meeting items.

1. Pre-Meeting Preparation

- a. ISBE recruitment of stakeholder panels
 - i. Recruited nine groups of 12-15 IAA stakeholders to convene for data review. The groups reviewed item data as follows:
 - Group 1- grade 3-5 Mathematics
 - Group 2 - grade 6-8 Mathematics
 - Group 3 - grade 11 Mathematics
 - Group 4 - grade 3-5 Reading
 - Group 5 - grade 6-8 Reading
 - Group 6 - grade 11 Reading
 - Group 7 - grade 4 Science
 - Group 8 - grade 7 Science
 - Group 9 - grade 11 Science.
- b. Pearson psychometricians prepared data related training and review material for meetings
 - i. Overview of IAA field test process
 - ii. Statistical review training materials
 - iii. Criteria for evaluating item statistics
 - iv. Supplemental field test results in tables and graphs
 - v. Field Test Statistics Review Books of item cards
 1. All field test statistics were provided in the Field Test Statistics Review Book, along with the content of each item:
 - a. Number of students who were administered the item—total and disaggregated by race and gender.

- b. Item difficulty,
 - c. Frequency of responses per rubric point,
 - d. Item to total student score correlations
 - e. Category averages
 - f. IRT step difficulties
 - g. IRT fit indices, and
 - h. DIF indicators.
 - 2. Additional field test statistics (e.g., reliability indices) and graphical aides to convey the field test results to the review panel were assembled by the psychometricians prior to the data review and used in the review as needed.
 - c. Pearson content specialist prepared IAA content related training and review material for meetings
 - i. Extended content frameworks
 - ii. Overview of item development and review process
 - iii. Rubrics
 - iv. Test administration materials
2. General Procedures during the Meeting
- a. Brief description of the test development process
 - b. Review of extended content frameworks
 - c. Review of test administration materials
 - d. Presentation of where data review fits in test development process
 - e. Panelists read a “Description of Item Statistics” handout and then write down any questions they may have.
 - i. A psychometrician described each of the item statistics in detail, explained their significance and answered any questions panelists had.
 - f. Proceeded through the Field Test Statistics Review book item by item. Facilitated panelists reading each item, reviewing the content alignment, and the associated statistics, and helped them make a determination as to whether each item is appropriate for use.
 - i. Potentially problematic items were discussed based on criteria mutually agreed upon by Pearson and ISBE. Our proposed statistical criteria were based on review of the FT data and included:
 - 1. Number of valid student responses per item must not be less than 75% of sample for form
 - 2. Item-total correlations must be equal to or greater than .40
 - 3. All item means must be at least two SEMs above 1.0 and below 4.0
 - 4. All items must have data for each rubric point
 - 5. Items with strong indication of DIF will trigger further scrutiny of item content and statistics in data review and bias review
 - 6. Items with unordered categories will be carefully reviewed

- ii. Discussed each item with the entire group to determine whether it should be retained or rejected. Reached group consensus.
3. Post-Meeting Tasks
 - a. After the data review meeting, Pearson reviewed rejected items with ISBE
 - b. Pearson and ISBE developed operational forms from retained items

Bias Review

Test Fairness

One aspect of the data review meetings was to assess potential bias based on DIF results and item content. Although bias in the items had been avoided through writer training and review processes, there is always the potential for bias to be detected through statistical analysis. It is important to include this step in the IAA development cycle because first and foremost, it is the responsibility of Pearson and ISBE to do everything possible to include only items on the IAA that will not: a) include content that is in anyway offensive or otherwise sensitive to minority students, b) discriminate against minority students, and c) provide inequitable test results for minority students. Pearson and ISBE do not intend to offend or disenfranchise minority students, parents, or teachers with an IAA item that is sensitive to a minority group if we have the means to prevent this condition. Likewise, Pearson and ISBE do not want to include an item that is biased in some way against a minority group, because the item may lead to inequitable test results.

Psychometric Issues

From a psychometric perspective, bias review did improve the quality of the IAA test by removing items that would potentially decrease test reliability and validity. Every item in the IAA pilot primarily measures proficiency in the content area it was written to assess, but to varying degrees each item also unintentionally measures other constructs that introduce unreliability to the item and the test. If one of the unintended constructs measured by an item is subgroup differences, and we can remove that item through a bias review, it is our responsibility to do so in order to increase the reliability and subsequent validity of the test. Bias in alternate assessments has received less research attention than has bias in general assessments and is therefore not as well understood as is bias in tests for the general student population. However, the psychometric standards are no lower for the alternate assessments than they are for general assessments and all psychometric procedures that are used to prevent bias in general assessments should be used to every degree possible for alternate assessments. A primary strength of the IAA events-based approach over the portfolio-based approach is that many more psychometric procedures can be usefully applied to the events-based tests—including DIF/DTF review by a bias panel.

Bias Review Process

As was described in the previous section, all IAA items were analyzed statistically for DIF using the pilot test data and complementary analysis methods. All items with any level of DIF identified were flagged and discussed by the data review panel.

Chapter V

Spring 2008 Operational Test Administration

Materials and training

Materials and teacher training for the IAA operational administration in Spring 2008 were nearly identical to those used in the Spring 2007 IAA writing administration described in the 2006-2007 IAA Technical Manual. Notable exceptions included the elimination of the teacher selected item (all items were mandatory for Spring 2008) and an increase in the specificity of items to decrease task ambiguity for teachers and increase standardization. Additional items were added to the Mathematics, Reading, and Science tests to increase content coverage. These changes were made by Pearson for ISBE based on feedback from the field and consultation with alternate assessment experts, and was explained in training and administration materials for teachers.

All training and administration materials for the IAA operational administration stressed choosing a time to test a student when it was best for that student and to administer the test in a familiar environment that is best for the student. Teachers were trained to administer the test to each student according to the student's primary mode of communication.

Administration

The Spring 2008 IAA was administered to grades 3-8, and 11 (Science only administered at grades 4, 7, and 11, and Writing only administered at grades 5, 6, 8, and 11). Guidelines for the operational IAA required the administration of the test by a certificated educator who may include but is not limited to the following:

- Teacher
- Administrator
- School psychologist
- Speech pathologist

Paraprofessionals were allowed to present the performance-based items to the student; however, a certificated educator must have observed and scored each item administered.

Each item was to be administered once, with sufficient time for a student to respond using his/her mode of communication, unless the item was interrupted and reasonably needed to be repeated from the beginning. Adequate wait time was defined as 3-5 seconds or longer (based on the student's mode of communication). If a student was having a rough day, teachers were provided the option of retesting the student at a later date within the 3-week testing window.

Test forms

Operational form design for the IAA was described in detail in the previous section. Each operational IAA form included five to ten scored items and one or two embedded field test items depending on subject. The operational items were selected from the pool of items surviving 2006

and/or 2007 data review meetings. One base operational form was administered per grade with unique field test items in each of seven versions of the base form.

Embedded Field Test Items

Embedded field test items (seven per grade for writing and 14 per grade for mathematics, reading, and science) were chosen from new items developed and reviewed for 2007-2008. Items were included in field test slots based on the needs of the item bank. Field test items were not included in calculating student scores. Results from analysis of the field test items are presented in Chapter VIII of this technical report.

Classical Item Analyses

Classical item analyses were conducted for each grade and subject of the Spring 2008 IAA Operational Administration as part of the 2008 operational calibration and equating. Analyses are reported both overall and by various ethnic and gender groups as sample size permitted.

The specific results of these analyses are presented in Appendix D. The first set of analyses provided the percentage of responses by item and item score for each grade and subject of the IAA. These results are presented in Tables D.1 – D.21.

Results from the second set of analyses provided information across the total test population regarding the score distributions for each item in the form of item means. Also, information regarding the degree to which each item is related to the total test score, and how well it may discriminate low from high scorers, is provided via the item total correlations. These results are presented in Tables D.22 – D.42 with item sequence and item content information.

Results from the third set of analyses focus on subgroups in the population and provide item means for each item and total raw score disaggregated by gender and ethnicity. These results are presented in Tables D.43 – D.63 with subgroup size information.

Reliability

Coefficient alpha reliability estimates were calculated for each IAA operational test by grade and subject. These results are in Table 5.1.

Table 5.1. Reliability estimate by grade and subject

Subject	Grade	Coefficient Alpha Reliability
Mathematics	3	0.92
	4	0.92
	5	0.90
	6	0.93
	7	0.90
	8	0.92
	11	0.90
Reading	3	0.93
	4	0.92
	5	0.92
	6	0.91
	7	0.92
	8	0.92
	11	0.94
Science	4	0.85
	7	0.83
	11	0.82
Writing	5	0.86
	6	0.85
	8	0.89
	11	0.88

The coefficient alpha estimates for the test form in each grade meets conventional guidelines for applied test reliability (i.e., $\alpha > .80$).

Calibration and Equating

Spring 2008 was the first operational administration of the IAA Mathematics, Reading, Science, and Grade 6 Writing tests. As such it was necessary to set the base IRT scale for these tests using the 2008 operational data. Pearson estimated IRT parameters for items appearing on these tests to establish the underlying theta scale for each category level of each item. These parameter estimates will serve to calibrate all students and test items onto the same underlying scale. The Rasch Partial Credit Model (RPCM) was selected because of its flexibility in accommodating a smaller n-count and for its ability to handle multiple-response category data. It also maintains a one-to-one relationship between derived scores (i.e., scale scores) and the underlying raw score scale. It is the underlying Rasch scale that facilitates equating of multiple test forms and allows for comparison of student performance across years. Additionally, the underlying Rasch scale facilitates the critical maintenance of equivalent performance standards across the years. The RPCM is defined via the following mathematical measurement model where, for a given item

involving m score categories, the probability of person n achieving score x on prompt i is given by:

$$P_{xni} = \frac{\exp \sum_{j=0}^x (B_n - D_{ij})}{\sum_{k=0}^{m_i} \exp \sum_{j=0}^k (B_n - D_{ij})},$$

where, $x = 0, 1, 2, \dots, m$, and,

$$\sum_{j=0}^0 (B_n - D_{ij}) \equiv 0$$

The RPCM provides the probability of a student scoring x on the m_i step of question/prompt i as a function of the student's proficiency level B_n (i.e., sometimes referred to as 'ability') and the step difficulties (D_{ij}) of the m steps in prompt i .

The Grade 5, 8, and 11 IAA Writing tests were administered for the first time in 2007 and the 'baseline' scale to which all future administrations will be equated was determined at that time. Spring 2008 represents the first year in which post-equating will be conducted under the IAA program.

The post-equating phase of the IAA base test was done in accordance with conventional common item procedures whereby the base test Rasch item difficulty values are compared with their previous calibrated values (from the 2007 live calibration) in order to derive a post-equating constant.

In Spring 2008 all items were included as part of the common item set for the three post-equated IAA writing assessments. The final post-equating constant was calculated as the difference in mean Rasch item difficulty of items in the common item set on the baseline (2007) scale versus the 2008 Rasch calibrated scale and can be represented as follows:

$$C = \bar{b}_{2003baseline} - \bar{b}_{2008calibrated}$$

where only those items remaining in the common item set after the stability check are used to compute the mean Rasch difficulty (\bar{b}). The post-equating adjustment constant is then added to the 2008 calibrated Rasch item difficulties for all items (not just those in the common item set) to bring them back onto the 2007 baseline scale.

The IAA equating design uses an iterative post-equating stability check procedure to eliminate from the calculation of the equating constant, test items whose Rasch item difficulty calibration differed more than expected from the pre-equated value. A post-equating constant was derived to put the calibration from the live administration onto the base scale. Using the items that determined the equating constant, the pre-equated item difficulty values were compared to the adjusted calibrated values from the live administration, and the item with the largest discrepancy was flagged. This item was a candidate for elimination from the set of items and the process of

determining a new equating constant if the discrepancy was larger than 0.50 logits. If, upon further review, it was decided that the item should be eliminated from the common item set then the equating constant was re-calculated, reapplied, and the stability check process repeated until all items with discrepancies larger than 0.50 logits had been considered. The stability check process was executed as follows:

1. Using the same method described above, calculate the provisional equating constant (C^*) using the remaining items in the equating set and apply it to the items in the test being equated.
2. Flag any items whose pre- and post- equated Rasch values differ by 0.50 logits or more.
3. Review the item with the largest absolute discrepancy from the equating set, and eliminate it if deemed appropriate.
4. Repeat steps 1 through 3 until no items have a discrepancy larger than 0.50 logits, until all such items have been reviewed and accepted, or until the minimum size of the equating set has been reached.
5. Examine the content representation of the items in the final equating set.

Evaluation of the sufficiency of the number and content representation of the final common item set should be evaluated as discussed in steps 4 and 5 above. Kolen and Brennan (2004) recommend that the common item set should be at least 20% of the test. Given the brevity of the IAA tests, content coverage was carefully considered when eliminating items. An item was retained if its elimination would result in a non-representative common item set.

IRT Analysis

The following IRT analyses were completed for all items administered in the Spring 2008 operational administration. Spring 2008 represents the first operational administration for Mathematics, Reading, Science, and Grade 6 Writing. IRT statistics presented for these tests were generated from the first operational IRT calibration described in the previous section of this Chapter. IRT statistics (i.e., step values and category averages) for Grade 5, 8, and 11 Writing are based on the 2008 post-equating described in the previous section of this Chapter.

- Item Fit Estimates. The extent to which the Rasch-Partial Credit model conformed to the data was estimated item by item. This diagnostic information was helpful in the selection of equating items eligible for psychometric equating.
 - INFIT MNSQ is the average of the INFIT mean-squares associated with the responses in each category. The expected values for all categories are 1.0.
- Category step values. These values are on the theta scale and represent the point on this scale at which one is more likely to receive the higher of two score point (e.g., the point on the ability continuum at which one is more likely to receive a score of 2 than a score of 1). This parameter indicates how difficult it is to observe a category, not how difficult it is to perform it.

- Average measure. The "average measure" for a category is the average ability of the people who respond in that category or to that distracter. Average measure = $\text{sum}(B_n - D_i) / \text{count of items in category}$. This is an empirical value. It is not a Rasch-model parameter.

From the IRT analysis, Pearson found that based on the average measure and category step values that the categories were properly ordered for all items, and that the information provided across categories for all items appropriately covered the -3 to +3 range of the theta scale. In Tables D.64 – D.84 the IRT results are presented by grade and subject.

IRT analysis also provides measures of test reliability. This reliability estimation method is based on the test information function (TIF) for each test and is conditional on the level of theta (i.e., test difficulty). As is presented in Figures D.1 – D.21, the TIF for each grade test differs across difficulty level and test score.

The point where test information is highest is also the point where the standard error of measurement (SEM) for the test is minimized, and the opposite is also true.

Scaling

Once standard setting for the new 2008 IAA operational tests (all math, reading, and science tests, and the grade 6 writing test) was complete and cut scores approved by the Illinois State Test Review Committee it was possible to derive scaling constants for these tests.

In keeping with work conducted in 2007 for the IAA Writing tests (based on Linacre, 2006) and ISBE's direction the scale score range was to extend from 30 to 70 with the Satisfactory performance level cut score anchored at 50. Attainable thetas were transformed to scale scores using the following formula:

$$SS = \text{Theta} * M1 + M2$$

M1 is calculated by dividing the distance between the bottom of the scale score range (30) and the scale score associated with the Satisfactory cut score (50) by the distance between theta value associated with the satisfactory cut score and the lowest attainable theta on the test. For example, the theta for the satisfactory cut score for grade 3 math is .64 and the lowest attainable theta on this test was -4.3171. Thus the equation for M1 would be written as follows:

$$M1 = (50 - 30) / (.64 - -4.3171)$$

M2 is calculated by computing the the difference between the scale score associated with the Satisfactory cut score (50) and the theta associated with this cut score (.64 for grade 3 math) and multiplying that figure by M1. This equation would appear as follows for grade 3 math:

$$M2 = (50 - .64) * M1$$

The transformation constants derived from this process and those calculated for the existing writing tests in 2007 are shown in the Tables 5.2 – 5.5.

Table 5.2. Mathematics Scaling Constants

Mathematics		
Grade	M1	M2
3	4.03	47.42
4	3.88	47.09
5	3.88	47.40
6	3.79	47.23
7	4.07	48.49
8	3.82	47.44
11	4.11	48.15

Table 5.3. Reading Scaling Constants

Reading		
Grade	M1	M2
3	3.56	45.62
4	3.55	45.60
5	3.55	45.63
6	3.78	46.64
7	3.63	46.33
8	3.59	46.63
11	3.58	45.56

Table 5.4. Science Scaling Constants

Science		
Grade	M1	M2
4	4.00	46.28
7	3.90	46.80
11	4.57	47.49

Table 5.5. Writing Scaling Constants

Writing		
Grade	M1	M2
5	3.74	46.71
6	4.23	48.10
8	4.44	47.20
11	4.42	48.58

Student score reports are generated in the Pearson scoring system based on raw to scale score conversion tables developed using the theta values from the Rasch calibration or equating of the events. This raw to scale score conversion table process for scoring student assessments is accepted best practices in the educational assessment industry and recommended by experts in the psychometric research arena (see Kolen & Brennan, 2004). One notable exception is 3pl pattern scoring and associated computer adaptive testing.

Chapter VI

Standard Setting for the IAA

Mathematics, Reading, Science, and Grade 6 Writing Tests

Introduction

On May 14 - 16, 2008 Pearson, under contract to the Illinois State Board of Education (ISBE), convened a panel of subject matter experts to set performance level cut scores on the Illinois Alternative Assessment (IAA) Mathematics test at grades 3-8 and 11, Reading test at grades 3-8 and 11, Science test at grades 4, 7, and 11, and Writing test at grade 6. Using a multiple method, iterative approach, including the Body of Work (Kingston, Kahl, Sweeney & Bay, 2001) method, item mapping (Mitzel, Lewis, Patz, & Green, 2001), and presentation of impact data in a Delphi process, the group reviewed the test and in relation to the Illinois Content Standards, recommended Foundational, Satisfactory, and Mastery cut scores. In arriving at each cut score, panelists considered group agreement and student impact. In this technical report the process and outcomes of the standard setting workshop are described.

Goal of the Standard Setting Panel

The goal of the meeting, as stated to the panelists, was to provide recommendations to ISBE on the appropriate placement of Foundational, Satisfactory, and Mastery performance level cut scores for the IAA Mathematics, Reading, Science, and Writing tests at the grades previously specified.

Overview of the Standard Setting Workshop

In this section, we provide an overview of the process for setting standards for the IAA. For each of the procedures described below, panelists were broken into five groups: Lower Reading (Grades 3 – 5), Upper Reading (Grade 6 – 8, & 11), Lower Mathematics (Grades 3 – 5), Upper Mathematics (Grade 6 – 8, & 11), and Science (Grades 4, 7, & 11). An additional committee, composed of panelists from the two Reading committees, was convened to complete an abbreviated interpolation process (detailed later in this chapter) to set standards for grade 6 Writing. In subsequent sections and appendixes, we provide additional details.

Body of Work Procedure

Pearson incorporated the Body of Work procedure (Kingston et al., 2001) to set performance standards on the IAA. In this approach, panelists review a set of student performances, or responses to prompts, that comprise a body of work. In the case of the IAA, panelists reviewed a set of videos developed to provide exemplars of student performances across the range of performance levels. The panelists were asked to respond to the videos, including agreement to the scores provided by the teachers in the videos. While reviewing each of the videos, the panelists are asked to keep in mind the performance level definitions (PLDs) developed by the ISBE (described below). In preparation for this round, panelists were trained on the various ways that a student could reach different levels of overall performance on the Body of Work that make up the IAA test, such as combinations of different levels of performance on each of the IAA tasks. At the conclusion of the Body of Work round, panelists were asked to independently set three cut scores (i.e., Foundational, Satisfactory, and Mastery) on the total test score based on

their review of the body of work. Once independent ratings were made, Pearson analyzed the individual cut score data and provided the aggregate and individual results back to the panelists. Group discussion was then facilitated in a Delphi process to encourage discussion of individual ratings.

Item mapping Standard Setting Process

Because the IAA has student performance evidence from both a Body of Work and an item-based perspective, Pearson incorporated the item mapping method (Mitzel et al., 2001) to setting performance standards as part of a multiple methods approach. This multiple method approach follows best practices in qualitative research (Denzin & Lincoln, 2007) for the use of multiple sources of evidence in judgmental tasks such as setting cut scores. For the item mapping method, panelists reviewed the entire test booklet, where items were ordered by difficulty on the theta scale, beginning with the easiest items. For the IAA, which uses constructed responses, items appeared multiple times in the ordered item booklet, once for each score point (i.e., 1 through 4). Panelists were asked to review the content of the items and compare each item score to the previous one keeping in mind why the preceding item might be easier and if this item score is measuring a higher level of performance. Next, panelists placed the bookmarks for each cut score set in the Body of Work round one in their test booklets. Finally, panelists were asked to consider the knowledge and skills a student must know to answer each item correctly or to attain each score point. Similar to the Body of Work method, panelists were asked to independently modify their cuts from round one using the information from the ordered item booklet; this time placing a bookmark on the page in the item ordered booklet where they thought the cut score needed to be set. Following the independent ratings, Pearson analyzed the individual bookmark placement data and once again provided aggregate and individual level statistics. As was done in round one, group discussion was facilitated in a Delphi process to encourage discussion of individual ratings.

Review of Impact Data

In addition to the round one Body of Work and round two item mapping steps, panelists were provided impact data based on their recommended cuts from the spring 2008 IAA operational administration. In the review of impact data, panelists were also shown the impact of the group median cut scores on percent of students within each performance level both in the total sample and by subgroup. Group discussion of the impact data was facilitated by informing panelists how this information was generated and how to incorporate it into the decisions that have been made up until this point. Panelists were then asked to independently use the impact data to make any adjustments to the cuts set in their ordered item booklet during round 3 of ratings.

Vertical Articulation

In the final stage of this multiple method approach to standard setting, panelists were asked to review the cuts across grades, within subject for consistency and reasonableness. In this step, panelists were presented with all of the cut scores across grades in a synthesis presentation. As part of this presentation, a subset of panelists across all grades was provided impact data for each cut score in each grade. Group discussion of vertical articulation focused on expected student progress across grades, given the student population, in a social moderation approach. As a group, panelists were asked to propose any potential adjustment to current scores based upon their knowledge of the student population and reasonable progression across grades. As

adjustments were suggested by individuals, the cut score data from the grade level groups was provided and rationales for keeping or adjusting cut scores were provided. The articulation committees were not allowed to move cut scores outside of the range of cut score ratings that existed in the grade level groups. The articulation committee was required to unanimously agree to a suggested change before it was implemented.

Performance Level Definitions (PLDs)

PLDs are a key element in most standard setting processes. PLDs define the content area knowledge, skills, and processes that examinees at a performance level are expected to possess and are based on inferences from their performance on a test. PLDs play a crucial role in all aspects of both the Body of Work and item mapping standard setting processes. In the Body of Work and item mapping procedures, panelists based their judgments on PLDs when they judged the probabilities of successful responses and placed their cut scores for the body of work and their item mappings in the ordered item booklet. The definitions of Entry, Foundational, Satisfactory, and Mastery performance that the ISBE developed and which the panelists used in recommending performance standards for the IAA are the public statements about what and how much Illinois educators expect students to know and be able to do. The progressing performance definitions defined aspiration goals for student achievement and targets for students and teachers.

IAA Standard Setting Workshop

In this section, we provide details of standard setting for the IAA tests.

Setting Three Performance Standards

For the IAA Mathematics, Reading, Science, and grade 6 Writing tests, cut scores were needed for three performance standards. Given that these IAA content areas were new for Illinois' students with disabilities population, identifying the cut scores presented a unique set of opportunities to the standard setting workshop panelists.

Feedback to Guide Panelist's Judgments

In the first round of the standard setting workshop, panelists were asked to provide a numerical value (on the raw score scale for the test under consideration) for their recommended cuts (panelists later placed these in the ordered item booklet). In each of the subsequent two rounds in the workshop, panelists placed a bookmark in an ordered item booklet to recommend a cut score between two performance levels. Panelists began rounds 2 and 3, and the vertical articulation round, with discussions about two types of feedback data. The feedback data were intended to increase panelists' understanding of their judgments about where they located their item mapping bookmarks and the impact of the performance standards recommended by the full panel.

Agreement data (i.e., group mean, median, and high and low recommended cut scores). This information was intended to guide panel discussions on individual panelist's judgments about item requirements and the placement of bookmarks and to facilitate convergence of judgments and recommended cut scores.

Impact data (i.e., the estimated percentage of students reaching the cut score on the operational test at each performance level). This information was provided to enable panelists to see the consequences of their recommended standards and possibly modify their judgments about items and cut scores accordingly. Impact data were reported based on the spring 2008 operational results. Impact data was not reviewed at the beginning of round 2, but was provided at the beginning of round 3 and prior to vertical articulation. Panelists received instructions on how to interpret the information and how to incorporate it into rounds 3 and vertical articulation judgments about cut score placement.

Standard Setting Panels

The demographic composition of the standard setting panel was critical to the process and politics of the standard setting process. Accordingly, ISBE recruited a standard setting panel of 51 panelists, primarily special and general education teachers, to serve across grades and subject areas. At the beginning of the workshops, the panelists were divided into five groups, Lower Reading (Grades 3 – 5), Upper Reading (Grade 6 – 8, & 11), Lower Mathematics (Grades 3 – 5), Upper Mathematics (Grade 6 – 8, & 11), and Science (Grades 4, 7, & 11), of nine to eleven panelists each. A committee, formed of reading committee panelists, was formed on site to conduct the Grade 6 Writing standard setting.

Recruiting Panel Members

ISBE, working with Pearson, conducted recruitment of the panel members. Pearson received a set of nominations from ISBE for potential panel recruitment and specifications for demographic composition of the panel. The goal of the recruitment process was to secure a total of 60 panelists that were representative of educators in the state of Illinois. Some of the panelists recruited for these meetings were ultimately unable to attend leading to the combined total of 51 panelists across committees.

Panel Composition

Upper Mathematics:

Ten panelists served on the upper grades Mathematics committee. The panel consisted of nine females and one male. Two of the panelists were Black and the other eight panelists were White. All but one panelist had a Master's degree; the remaining panelist had a Bachelor's degree. Nine of the panelists indicated that they were teachers and one did not indicate a job title. The majority of the committee members were special education teachers.

Lower Mathematics:

Eleven panelists served on the lower grades Mathematics committee. The panel consisted of ten females and one male. One of the panelists was of Native American descent and the other ten panelists were White. All but one panelist had a Master's degree; the remaining panelist had a Bachelor's degree. Eight of the panelists indicated that they were teachers and three did not indicate a job title. The majority of the committee members were special education teachers.

Upper Reading:

Eleven panelists served on the upper grades Reading committee. The panel consisted of ten females and one male. One of the panelists was Black, one was Hispanic, and the other ten panelists were White. Seven of the panelists had a Master's degree; the remaining four panelists had a Bachelor's degree. Eight of the panelists indicated that they were teachers and three did not indicate a job title. The majority of the committee members were special education teachers.

Lower Reading:

Ten panelists served on the lower grades Reading committee. The panel was all female. All of the panelists were White. Eight of the panelists had a Master's degree, one panelist had a Bachelor's degree, and one indicated "other" in this field. Five of the panelists indicate that they

were teachers, one listed language specialist, one was a program coordinator, and three did not indicate a job title. The majority of the committee members were special education teachers.

Science:

Nine panelists served on the Science committee. The panel consisted of seven females and two males. All of the panelists were White. Two of the panelists had a Doctorate, five had a Master's degree, and two panelists had a Bachelor's degree. Eight of the panelists indicated that they were teachers and one did not indicate a job title. The majority of the committee members were special education teachers.

Writing:

The writing committee was composed of nine members of the upper and lower grades Reading committees that were not needed for vertical articulation.

Workshop Procedures

Training

The session began with a review of the purpose and agenda, followed by panelist training, which is an essential element of a standard setting workshop. Panelists received training on the multiple method approach implemented here for standard setting, the Illinois Learning Standards, the test specifications, and the Performance Level Definitions (PLDs) for each performance standard. They watched videos of students taking the test and the subsequent teacher ratings of performance. Panelists were also provided information regarding the scoring procedures, scaling procedures, and other details of the testing process that were necessary for recommending performance standards. They learned about the meaning of the threshold student in placing their cut scores.

The training was organized into five parts:

1. General overview of setting performance standards on student assessments
2. Orientation to the Body of Work standard setting procedure
3. Orientation to the item mapping standard setting procedure
4. Orientation to the IAA test materials and training process
5. Orientation to Illinois Learning Standards and Performance Level Definitions

Pearson staff led the training of the panelists. The Pearson Workshop Leader trained the panelists on using the Body of Work, Item mapping, and content standards and other materials. The training session is discussed below.

General overview of setting performance standards on student assessments

Panelists were provided an overview of standard setting, including a description of the standard setting process, the importance of standard setting, and some basic vocabulary to familiarize panelists with general procedures. The importance of the standard setting workshop and the panelist judgments was emphasized throughout the overview. Panelists were encouraged to ask questions and discuss the standard setting process in general and specific to the IAA.

Orientation to the Body of Work standard setting procedure

Next, panelists were trained on the Body of Work approach. In this section, panelists were presented with the rubric for scoring, scoring of the constructed response items, how the overall score is computed, and provided examples of the test being administered to students receiving certain item scores. Several examples of how discrete student performances on IAA items could combine to form the larger IAA Body of Work were discussed. The panelists were provided insights on how to extrapolate from the set of student performance videos used in the workshop to all possible combinations of student performance.

For establishing their cut scores using the Body of Work method, panelists received the following recommendations for making their cuts:

1. Review the IAA PLDs
2. Review the videos representing a range of student IAA performance
3. Review the rubrics

Panelists were encouraged to ask questions and discuss the how the Body of Work standard setting process would be applied to the IAA.

Orientation to the item mapping standard setting procedure

Panelists were next trained on the item mapping procedure for standard setting. The panelists learned that the item mapping method is a procedure for setting performance standards that has been used in more than 20 states and has withstood legal challenges. Panelists were trained on the ordered item booklet and how the booklet was assembled by placing each of the four score points for each item in order of difficulty on the theta scale. Several examples and graphics were presented to the panelists to accentuate learning of this concept and how it was important to the IAA standard setting tasks. As a part of this section in the training panelists were presented with the threshold student concept. Panelists provided examples of threshold students from their own experiences in the teaching and assessing students in the SWD population. Panelists were asked to use this concept in deciding the placement of the bookmark.

The panelists received the following instructions for placing a bookmark.

1. The panelists read each item in the ordered item booklet and identified the knowledge and skills required to respond successfully to the item.
2. The panelists reviewed the Performance Level Definitions.
3. Panelists were told that they were to find the point where a student on the threshold of a performance level as described by the PLDs (e.g., a borderline Foundational, Satisfactory, or Mastery) would have a 50% probability of answering the last item correctly, but less than a 50% probability of answering the next item in the booklet correctly.
4. Finally, the panelists found the location in the ordered item booklet that separated groups of students into performance level categories and then placed a bookmark on the page presenting the last item/score point a threshold student would be more likely than not to achieve successfully.

Panelists had a chance to review the steps and to ask questions as we modeled the steps for them using a practice ordered item booklet. The Pearson staff led the panelists through all items as a

group, discussing what was required to answer the item correctly and what made the item difficult. Finally, we compared the items with a Performance Level Definition (in the relevant grade and content area) to determine where the bookmark would be placed to separate the items a student should answer correctly 50% of the time and those a student would answer correctly less than 50% of the time. Panelists were instructed to think about the item booklet as a different source of information than that presented in the Body of Work round, while considering that they needed to synthesize this information to successfully complete their task. Finally, we trained the panelists on how to use an item map and a rating form, instructing them to place the bookmark on the last item the student *should* answer correctly and record that page number on the rating form.

Orientation to the IAA test materials and training process

Panelists were given materials on the test specifications and an explanation of how we developed the pool of items from these standards. These specifications included a general description of the assessment, test booklets, answer sheets, administration manuals, and scoring rubrics. Panelists were instructed to use these documents to familiarize themselves with the content standards, how the test was designed, and what students were specifically expected to know. Discussion of the materials was facilitated, with panelists who had participated in IAA test development and/or administered the IAA providing subject matter expertise and relevant personal experiences with the test.

Orientation to Illinois content standards and Performance Level Definitions

The panelists were trained on the relationships between the Illinois Learning Standards, the Illinois Alternate Assessment Frameworks, and the IAA items. Pearson staff explained how the general education learning standards were extended to the alternate assessment frameworks through the essence of each standard and that the integrity of the standards across the population and across subjects and grades remained intact. Further, the panelists were trained on how the IAA items were written to tap into the writing construct for each grade as defined by the learning standards. Panelists with experience in the IAA item writing and review activities provided insights to other panelists and assisted in describing the alternate assessment frameworks.

Pearson then introduced the IAA Performance Level Definitions defining what students should be able to do at the Entry, Foundational, Satisfactory, and Mastery levels. Pearson gave the panelists time to discuss the interpretation and implications among themselves and to familiarize themselves with the definitions. Panelists were led through a process of creating behavioral anchors to provide enhancements to the PLDs for better understanding. Each group described students from the students with disabilities (SWD) population who would be within the performance level described by each PLD and found agreement on what the students within that level would be able to know and do.

Standard Setting

For the standard setting workshop, panelists were divided into five groups: Lower Reading (Grades 3 – 5), Upper Reading (Grade 6 – 8, & 11), Lower Mathematics (Grades 3 – 5), Upper Mathematics (Grade 6 – 8, & 11), and Science (Grades 4, 7, & 11). Each group was led by a member of the Pearson Psychometric Services staff trained and experienced in facilitation of the standard setting process. Each of these committees was to setting standards for three or four

grades within a subject area. The committees began by setting standards for the lowest grade in their range using the full three-round process, set standards for highest grade in their range using the full three-round process, and then completed a one-round validation of bookmark placements derived for the middle grade(s) in the range from a linear interpolation of the cut scores set for the highest and lowest grades. On the final day of standard setting, an articulation committee was convened for each subject area to address the appropriateness of cut score recommendations across all grades. These committees, with the exception of Science, enlisted a subset of the members of the lower and upper grade committees for the subject of interest. Because Science only required one committee, the full committee was retained for the articulation process. A sixth committee, composed of panelists from the upper and lower reading committees not involved in the articulation process, met on the final day to complete a one-round validation of grade 6 Writing cut scores derived from interpolation of the grade 5, 8, and 11 Writing cut scores set in 2007. The standard setting is detailed below.

Conducting the Body of Work Method

Prior to the Body of Work round, panelists were provided the rating forms that they would be using to record their cuts and to provide rationales. Panelists were asked to review the materials and ask any questions that they might have. Group facilitators re-presented the Body of Work training materials to the panelists to ensure adequate understanding of the process. Panelists were then presented with a committee-specific set of videos showing students earning different score points completing the full IAA assessment for that subject. Using the PLDs and keeping the videos presented in mind, panelists were asked to independently set three cuts on the IAA raw score scale, record the raw score associated with these cuts on their recording sheet, and to provide a rationale for each of their cuts. The group facilitator collected the ratings and computed the mean, median, minimum, and maximum. This information was then presented back to the panel and a discussion of the cuts and their rationales ensued using a Delphi focus group method. This concluded round 1 of standard setting. Round 1 recommendations are reported by grade and subject in Tables 6.1 – 6.13. Please note that the round 1 recommendations are on the raw score metric.

Table 6.1. Round 1 Cut Score Recommendations – Grade 3 Mathematics

	Foundational	Satisfactory	Mastery
Mean	16.00	24.73	34.18
Median	16.00	25.00	34.00
Minimum	15.00	23.00	31.00
Maximum	17.00	26.00	37.00

Table 6.3. Round 1 Cut Score Recommendations – Grade 5 Mathematics

	Foundational	Satisfactory	Mastery
Mean	16.64	25.36	34.27
Median	17.00	26.00	34.00
Minimum	15.00	23.00	31.00
Maximum	18.00	26.00	37.00

Table 6.4. Round 1 Cut Score Recommendations – Grade 6 Mathematics

	Foundational	Satisfactory	Mastery
--	--------------	--------------	---------

2007-2008 IAA Annual Technical Manual APPENDICES

Mean	14.80	23.80	32.60
Median	16.00	24.50	32.50
Minimum	8.00	18.00	29.00
Maximum	19.00	26.00	35.00

Table 6.7. Round 1 Cut Score Recommendations – Grade 11 Mathematics

	Foundational	Satisfactory	Mastery
Mean	14.60	23.80	32.70
Median	15.50	24.50	32.50
Minimum	10.00	19.00	30.00
Maximum	17.00	26.00	35.00

Table 6.8. Round 1 Cut Score Recommendations – Grade 3 Reading

	Foundational	Satisfactory	Mastery
Mean	11.80	20.00	28.70
Median	11.00	20.00	28.00
Minimum	10.00	15.00	28.00
Maximum	15.00	23.00	32.00

Table 6.9. Round 1 Cut Score Recommendations – Grade 5 Reading

	Foundational	Satisfactory	Mastery
Mean	11.70	19.90	28.90
Median	11.50	20.50	28.00
Minimum	10.00	13.00	27.00
Maximum	15.00	23.00	32.00

Table 6.10. Round 1 Cut Score Recommendations – Grade 6 Reading

	Foundational	Satisfactory	Mastery
Mean	12.36	21.09	30.27
Median	12.00	21.00	30.00
Minimum	10.00	20.00	30.00
Maximum	15.00	23.00	31.00

Table 6.11. Round 1 Cut Score Recommendations – Grade 11 Reading

	Foundational	Satisfactory	Mastery
Mean	11.27	20.45	29.09
Median	11.00	20.00	29.00
Minimum	10.00	20.00	28.00
Maximum	13.00	23.00	31.00

Table 6.12. Round 1 Cut Score Recommendations – Grade 4 Science

	Foundational	Satisfactory	Mastery
Mean	8.89	14.78	19.44
Median	9.00	15.00	20.00
Minimum	8.00	13.00	18.00
Maximum	10.00	16.00	20.00

Table 6.13. Round 1 Cut Score Recommendations – Grade 11 Science

	Foundational	Satisfactory	Mastery
Mean	8.67	15.00	19.56
Median	9.00	15.00	19.00
Minimum	8.00	14.00	19.00
Maximum	9.00	16.00	21.00

Reviewing the Ordered Item Booklet & Conducting the Item Mapping Process

Ordered item booklets were provided to the panelists, who were asked to make a judgment about “the divide between items that a student at the threshold of a performance level (the minimally qualified student) should master from those items that are not necessary to master” (Mitzel et al., 2001, p. 254). The panelists were asked to place a bookmark at that page of the ordered item booklet. Prior to making their ratings, panelists were asked to place bookmarks on the three pages that corresponded to their first round (Body of Work) cut scores. Panelists were presented a portion of the broader training regarding the ordered item booklet and the item mapping procedure until all panelists understood the procedure.

In the item mapping standard setting, score points from each item on the spring 2008 operational form--representative of the range of content and difficulty of the item bank--were rank ordered according to their level of Rasch difficulty using the average total-test-based theta estimate associated with the students who endorsed each score point. The difficulty estimates were based on operational test data obtained from the spring 2008 operational test. The ordered item maps for each test are presented in Appendix E.

Using the ordered item booklet and keeping the first round cut scores in mind, panelists were asked to independently make three cuts by adjusting the placement of the bookmarks in the ordered item booklet, write these cuts on their recording sheet, and to provide rationale for each of their cuts (or adjustment of cuts). Similar to the Body of Work, the group facilitator collected the ratings and computed the mean, median, minimum, and maximum. This information was then presented back to the panel and a discussion of the cuts and their rationale ensued. This concluded round 2 of standard setting. Round 2 recommendations are reported by grade and subject in Tables 6.14 – 6.23. Please note that the round 2 recommendations are expressed in terms of ordered item booklet page number.

Table 6.14. Round 2 Cut Score Recommendations – Grade 3 Mathematics

	Foundational	Satisfactory	Mastery
Mean	11.36	20.36	30.36
Median	13.00	21.00	31.00
Minimum	4.00	15.00	25.00
Maximum	16.00	24.00	34.00

Table 6.15. Round 2 Cut Score Recommendations – Grade 5 Mathematics

	Foundational	Satisfactory	Mastery
Mean	13.18	21.36	31.64
Median	13.00	21.00	32.00

Minimum	11.00	16.00	26.00
Maximum	16.00	25.00	34.00

Table 6.16. Round 2 Cut Score Recommendations – Grade 6 Mathematics

	Foundational	Satisfactory	Mastery
Mean	9.50	16.50	28.10
Median	9.50	16.00	26.50
Minimum	7.00	12.00	24.00
Maximum	13.00	22.00	33.00

Table 6.17. Round 2 Cut Score Recommendations – Grade 11 Mathematics

	Foundational	Satisfactory	Mastery
Mean	11.30	20.60	30.90
Median	11.00	21.00	30.50
Minimum	8.00	13.00	27.00
Maximum	16.00	29.00	37.00

Table 6.18. Round 2 Cut Score Recommendations – Grade 3 Reading

	Foundational	Satisfactory	Mastery
Mean	10.90	18.50	28.00
Median	10.00	20.00	28.00
Minimum	6.00	9.00	23.00
Maximum	18.00	24.00	31.00

Table 6.19. Round 2 Cut Score Recommendations – Grade 5 Reading

	Foundational	Satisfactory	Mastery
Mean	15.30	22.80	30.40
Median	17.00	23.50	31.50
Minimum	9.00	18.00	26.00
Maximum	19.00	25.00	33.00

Table 6.20. Round 2 Cut Score Recommendations – Grade 6 Reading

	Foundational	Satisfactory	Mastery
Mean	9.18	17.45	26.73
Median	10.00	17.00	27.00
Minimum	4.00	12.00	22.00
Maximum	14.00	23.00	30.00

Table 6.21. Round 2 Cut Score Recommendations – Grade 11 Reading

	Foundational	Satisfactory	Mastery
Mean	12.36	22.00	30.36
Median	13.00	22.00	30.00
Minimum	9.00	18.00	27.00
Maximum	15.00	25.00	33.00

Table 6.22. Round 2 Cut Score Recommendations – Grade 4 Science

	Foundational	Satisfactory	Mastery
Mean	6.00	11.11	17.00
Median	6.00	11.00	17.00
Minimum	4.00	9.00	15.00
Maximum	8.00	15.00	19.00

Table 6.23. Round 2 Cut Score Recommendations – Grade 11 Science

	Foundational	Satisfactory	Mastery
Mean	6.33	11.89	17.22
Median	6.00	12.00	16.00
Minimum	5.00	10.00	15.00
Maximum	8.00	14.00	20.00

Presentation of Impact Data

Prior to beginning round 3, Pearson staff computed the impact data for the median page number cuts recommended by the panelists following round 2. The impact data was computed by using the corresponding theta value for a particular cut score and computing the number of students that would pass that cut. This data was based on the spring 2008 operational test calibration sample that was used to estimate item parameters for the test. Round 3 began with a presentation of this impact data. Discussion focused on the recommended cut scores from round 2 and the items between the high and low cut scores. In addition, discussion included a consideration of the degree to which panelists' cut scores had or had not converged since round 1 as a means of discussing the degree to which panelists agreed about the difficulty of items and the influence of the impact data.

As before, panelists were again to place (adjust) their bookmarks one last time. Panelists recorded their cuts on the recording sheet and entered their rationale for their final cut score. The group facilitator collected the ratings and computed, the mean, median, minimum, and maximum. This information was then presented back to the panel and a discussion of the cuts and their rationale ensued. Panelists were then presented with the impact of their final cut score recommendations for discussion. This concluded the third, and final, round of standard setting. Round 3 recommendations are reported by grade and subject in Tables 6.24 – 6.33. Please note that the round 3 recommendations are expressed in terms of ordered item booklet page number.

Table 6.24. Round 3 Cut Score Recommendations – Grade 3 Mathematics

	Foundational	Satisfactory	Mastery
Mean	13.18	22.00	31.82
Median	13.00	21.00	33.00
Minimum	11.00	19.00	28.00
Maximum	15.00	24.00	34.00

Table 6.25. Round 3 Cut Score Recommendations – Grade 5 Mathematics

	Foundational	Satisfactory	Mastery
Mean	13.36	21.91	32.18
Median	13.00	21.00	32.00
Minimum	12.00	19.00	30.00
Maximum	15.00	25.00	34.00

Table 6.26. Round 3 Cut Score Recommendations – Grade 6 Mathematics

	Foundational	Satisfactory	Mastery
Mean	11.90	20.70	34.10
Median	12.00	21.00	34.00
Minimum	11.00	20.00	34.00
Maximum	12.00	21.00	35.00

Table 6.27. Round 3 Cut Score Recommendations – Grade 11 Mathematics

	Foundational	Satisfactory	Mastery
Mean	11.10	19.00	31.10
Median	11.00	19.00	31.00
Minimum	10.00	19.00	31.00
Maximum	13.00	19.00	32.00

Table 6.28. Round 3 Cut Score Recommendations – Grade 3 Reading

	Foundational	Satisfactory	Mastery
Mean	15.20	22.40	30.60
Median	15.00	22.50	31.00
Minimum	13.00	19.00	29.00
Maximum	18.00	24.00	31.00

Table 6.29. Round 3 Cut Score Recommendations – Grade 5 Reading

	Foundational	Satisfactory	Mastery
Mean	16.70	23.80	31.40
Median	17.00	24.00	32.00
Minimum	15.00	23.00	27.00
Maximum	18.00	24.00	33.00

Table 6.30. Round 3 Cut Score Recommendations – Grade 6 Reading

	Foundational	Satisfactory	Mastery
Mean	11.73	20.18	30.00
Median	12.00	21.00	30.00
Minimum	10.00	17.00	27.00
Maximum	13.00	22.00	32.00

Table 6.31. Round 3 Cut Score Recommendations – Grade 11 Reading

	Foundational	Satisfactory	Mastery
Mean	12.73	22.73	30.91
Median	13.00	22.00	31.00
Minimum	9.00	21.00	30.00
Maximum	14.00	25.00	32.00

Table 6.32. Round 3 Cut Score Recommendations – Grade 4 Science

	Foundational	Satisfactory	Mastery
Mean	6.44	12.33	18.78
Median	7.00	13.00	19.00
Minimum	4.00	10.00	17.00
Maximum	8.00	14.00	23.00

Table 6.33. Round 3 Cut Score Recommendations – Grade 11 Science

	Foundational	Satisfactory	Mastery
Mean	6.67	12.56	17.89
Median	6.00	12.00	18.00
Minimum	6.00	10.00	16.00
Maximum	8.00	15.00	20.00

As mentioned previously, each committee completed this process beginning with the lowest grade their committee would consider. The process was then completed for the highest grade the committee would consider. For the middle grade(s) in each committee the initial cut scores were interpolated using linear regression based on the median cut scores for the highest and lowest grade after Round 3. These interpolated cut scores were presented to the panelists who then reviewed item content and difficulty and facilitators led the panelists through a discussion of the appropriateness of the interpolated cut scores based on the item content information. Impact data based on the interpolated cut scores was also presented and discussed by the panelists. After these discussions panelists were given the opportunity to individually modify the cut score through one round of item mapping.

This process was also used to set the cut scores for Grade 6 Writing. The cut scores for Grade 5, 8, and 11 Writing were established in 2007 in order to provide scores for the 2007 operational administrations. In order to ensure that the Grade 6 cut scores were not set without the context of that meeting the interpolation process was used to provide a starting point. For Writing, the interpolation was based on the standards set for Grade 5, 8, and 11 Writing in 2007. Post interpolation cut score recommendations are presented in Tables 6.34 – 6.41. Please note that these recommendations are expressed in terms of ordered item booklet page number.

Table 6.34. Post Interpolation Cut Score Recommendations – Grade 4 Mathematics

	Foundational	Satisfactory	Mastery
Mean	12.18	20.36	32.00
Median	12.00	20.00	33.00
Minimum	11.00	19.00	28.00

2007-2008 IAA Annual Technical Manual APPENDICES

Maximum	13.00	23.00	33.00
---------	-------	-------	-------

Table 6.35. Post Interpolation Cut Score Recommendations – Grade 7 Mathematics

	Foundational	Satisfactory	Mastery
Mean	12.18	20.36	32.00
Median	12.00	20.00	33.00
Minimum	11.00	19.00	28.00
Maximum	13.00	23.00	33.00

Table 6.36. Post Interpolation Cut Score Recommendations – Grade 8 Mathematics

	Foundational	Satisfactory	Mastery
Mean	11.22	20.89	29.44
Median	11.00	21.00	29.00
Minimum	11.00	20.00	29.00
Maximum	13.00	21.00	33.00

Table 6.37. Post Interpolation Cut Score Recommendations – Grade 4 Reading

	Foundational	Satisfactory	Mastery
Mean	14.67	23.67	33.67
Median	15.00	23.00	34.00
Minimum	12.00	22.00	31.00
Maximum	15.00	26.00	35.00

Table 6.38. Post Interpolation Cut Score Recommendations – Grade 7 Reading

	Foundational	Satisfactory	Mastery
Mean	13.00	20.80	28.90
Median	13.00	21.00	29.00
Minimum	13.00	20.00	28.00
Maximum	13.00	21.00	29.00

Table 6.39. Post Interpolation Cut Score Recommendations – Grade 8 Reading

	Foundational	Satisfactory	Mastery
Mean	12.50	20.00	29.00
Median	12.50	20.00	29.00
Minimum	12.00	20.00	29.00
Maximum	13.00	20.00	29.00

Table 6.40. Post Interpolation Cut Score Recommendations – Grade 7 Science

	Foundational	Satisfactory	Mastery
Mean	7.00	13.33	19.11
Median	7.00	13.00	19.00
Minimum	7.00	13.00	19.00
Maximum	7.00	15.00	20.00

Table 6.41. Post Interpolation Cut Score Recommendations – Grade 6 Writing

	Foundational	Satisfactory	Mastery
Mean	5.89	10.67	17.56
Median	6.00	10.00	18.00
Minimum	4.00	9.00	14.00
Maximum	8.00	13.00	19.00

Evaluation of Standard Setting Workshop

After standard setting was complete, but prior to vertical articulation panelists, were provided the opportunity to give evaluative feedback concerning the process and outcomes of the standard setting workshop through a standardized evaluation form (see Appendix F). The evaluation form included both closed- and open-ended response items. Aggregated results of the closed-ended items are presented in Appendix G. Overall, the evaluation results indicate the panelists were satisfied and understood the standard setting process. They felt that the workshop was useful and group facilitators were effective. Panelists seemed to agree that the multiple method approach implemented here was effective for setting standards on the IAA.

Vertical Articulation

Final cut scores for each grade were presented as a set to a subject specific articulation committee for a vertical articulation round. In this round, panelists were asked to consider impact trends across grades, and, if necessary, modify cut scores in order to ensure that the standards better reflect panelists' judgments regarding student expectations across grades. Any changes in cut scores during this meeting were not allowed to move outside of the range of individual panelist ratings during the final round of standard setting. In order for a cut score to be moved by the articulation committee complete consensus was required. The Mathematics group did not modify any items during articulation. The Reading group agreed to modify only one cut score: the Satisfactory cut score for sixth grade. The Science group agreed to move three cut scores: the foundational cut score for eleventh grade and the Mastery cut scores for fourth and seventh grades were all raised one page. Articulation was not conducted for Grade 6 Writing as the cut scores for the other writing tests were already approved and could not be altered.

Recommended Final Cut Scores and Impact Data

The ultimate outcomes of the standard setting activity were the recommended cut scores. In this section the recommended cut scores after articulation for each of the performance levels are presented for Mathematics, Reading, and Science in Tables 6.42 – 6.44 respectively. The estimated impact each performance standard would have on subgroups and the total IAA population is presented by grade for Mathematics, Reading, and Science in Tables 6.45 – 6.61 (impact indicates the percentage of students meeting or exceeding each performance level). Final recommended cut scores and impact data for grade 6 Writing are presented in Table 6.62.

Table 6.42. Mathematics Foundational, Satisfactory, and Mastery Final Cut Scores

Cut Score by Ordered Item Booklet Page							
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Foundational	13	12	13	12	11	11	11
Satisfactory	21	20	21	21	19	21	19

Mastery	33	33	32	34	31	33	31
---------	----	----	----	----	----	----	----

Table 6.43. Reading Foundational, Satisfactory, and Mastery Final Cut Scores

Cut Score by Ordered Item Booklet Page							
	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 11
Foundational	15	15	17	12	13	13	13
Satisfactory	23	23	23	21	21	20	22
Mastery	31	34	32	30	29	29	31

Table 6.44. Science Foundational, Satisfactory, and Mastery Final Cut Scores

Cut Score by Ordered Item Booklet Page			
	Grade 4	Grade 7	Grade 11
Foundational	7	7	7
Satisfactory	13	13	12
Mastery	20	20	18

Table 6.45. Impact of Final Cut Scores – Grade 3 Math

	All	Male	Female	Black	Hispanic	White
Foundational	79.04	76.83	80.22	71.81	75.66	82.71
Satisfactory	64.88	65.60	64.50	57.53	59.73	68.87
Mastery	30.08	30.50	29.85	25.87	28.32	32.93

Table 6.46. Impact of Final Cut Scores – Grade 4 Math

	All	Male	Female	Black	Hispanic	White
Foundational	83.92	80.58	85.70	81.76	83.56	85.18
Satisfactory	68.07	64.09	70.18	69.93	63.56	68.69
Mastery	31.72	28.18	33.59	33.11	28.89	32.22

Table 6.47. Impact of Final Cut Scores – Grade 5 Math

	All	Male	Female	Black	Hispanic	White
Foundational	82.01	82.10	81.97	79.51	78.48	84.45
Satisfactory	64.85	63.37	65.69	57.29	66.82	67.67
Mastery	26.34	25.93	26.58	24.31	27.35	26.88

Table 6.48. Impact of Final Cut Scores – Grade 6 Math

	All	Male	Female	Black	Hispanic	White
Foundational	85.09	81.99	86.79	85.03	84.52	85.82
Satisfactory	70.71	67.92	72.24	67.80	64.68	75.06
Mastery	26.50	25.33	27.14	22.88	24.60	29.22

Table 6.49. Impact of Final Cut Scores – Grade 7 Math

	All	Male	Female	Black	Hispanic	White
Foundational	79.95	80.68	79.51	72.46	77.35	83.95
Satisfactory	64.99	67.62	63.41	57.78	64.46	67.65
Mastery	23.94	23.61	24.14	20.06	22.30	26.19

Table 6.50. Impact of Final Cut Scores – Grade 8 Math

	All	Male	Female	Black	Hispanic	White
Foundational	86.47	87.00	86.13	82.09	86.07	88.93
Satisfactory	68.82	68.67	68.92	65.17	68.03	70.36

Mastery	30.92	31.33	30.65	29.35	30.33	32.35
---------	-------	-------	-------	-------	-------	-------

Table 6.51. Impact of Final Cut Scores – Grade 11 Math

	All	Male	Female	Black	Hispanic	White
Foundational	82.33	83.60	81.53	79.09	81.21	84.58
Satisfactory	70.06	68.76	70.88	65.51	63.76	73.50
Mastery	33.94	35.73	32.81	31.36	25.50	36.38

Table 6.52. Impact of Final Cut Scores – Grade 3 Reading

	All	Male	Female	Black	Hispanic	White
Foundational	81.09	80.55	81.37	74.52	74.45	85.01
Satisfactory	60.49	61.33	60.05	52.90	53.30	65.07
Mastery	28.73	29.98	28.06	23.55	24.23	32.08

Table 6.53. Impact of Final Cut Scores – Grade 4 Reading

	All	Male	Female	Black	Hispanic	White
Foundational	78.42	78.29	78.49	75.00	78.22	79.38
Satisfactory	60.17	58.25	61.20	60.14	57.33	60.18
Mastery	22.01	19.62	23.28	22.30	18.67	22.68

Table 6.54. Impact of Final Cut Scores – Grade 5 Reading

	All	Male	Female	Black	Hispanic	White
Foundational	74.83	76.75	73.74	68.06	73.09	78.55
Satisfactory	59.52	61.32	58.50	53.47	58.30	62.98
Mastery	21.96	24.49	20.52	19.44	23.77	23.50

Table 6.55. Impact of Final Cut Scores – Grade 6 Reading

	All	Male	Female	Black	Hispanic	White
Foundational	82.01	78.80	83.78	80.51	76.10	85.57
Satisfactory	60.89	59.47	61.67	55.08	53.39	66.38
Mastery	31.91	31.14	32.33	26.84	27.49	36.19

Table 6.56. Impact of Final Cut Scores – Grade 7 Reading

	All	Male	Female	Black	Hispanic	White
Foundational	82.25	84.11	81.14	76.05	81.53	85.38
Satisfactory	61.96	65.18	60.02	55.99	56.79	66.28
Mastery	28.20	31.07	26.47	23.65	26.83	30.64

Table 6.57. Impact of Final Cut Scores – Grade 8 Reading

	All	Male	Female	Black	Hispanic	White
Foundational	82.63	83.81	81.87	77.61	80.08	85.96
Satisfactory	66.10	65.61	66.42	61.94	64.63	68.47
Mastery	29.65	30.22	29.29	26.12	30.89	30.79

Table 6.58. Impact of Final Cut Scores – Grade 11 Reading

	All	Male	Female	Black	Hispanic	White
Foundational	86.88	87.67	86.38	83.62	83.89	88.96

2007-2008 IAA Annual Technical Manual APPENDICES

Satisfactory	69.42	69.28	69.50	64.11	65.10	72.84
Mastery	29.89	31.17	29.08	26.83	22.82	32.84

Table 6.59. Impact of Final Cut Scores – Grade 4 Science

	All	Male	Female	Black	Hispanic	White
Foundational	85.55	83.65	86.56	83.33	82.59	87.10
Satisfactory	66.88	65.83	67.44	68.37	64.29	67.10
Mastery	28.03	27.67	28.22	25.51	27.23	28.90

Table 6.60. Impact of Final Cut Scores – Grade 7 Science

	All	Male	Female	Black	Hispanic	White
Foundational	86.83	87.97	86.14	82.28	84.32	89.70
Satisfactory	67.07	70.20	65.20	63.66	60.63	71.04
Mastery	40.52	39.86	40.92	36.64	35.19	44.53

Table 6.61. Impact of Final Cut Scores – Grade 11 Science

	All	Male	Female	Black	Hispanic	White
Foundational	87.62	87.39	87.77	86.36	84.56	89.06
Satisfactory	73.15	70.95	74.54	67.13	66.44	77.66
Mastery	33.22	28.83	35.99	30.77	30.87	35.23

Table 6.62. Final Cut Scores and Associated Impact – Grade 6 Writing

	Page	All	Male	Female	Black	Hispanic	White
Foundational	6	74.55	73.12	75.34	72.03	67.86	78.46
Satisfactory	10	69.22	68.23	69.76	68.64	61.51	72.95
Mastery	18	24.18	25.56	23.43	22.88	22.22	25.09

Conclusions

Pearson endorses these cut score recommendations for the IAA Mathematics, Reading, Science, and grade 6 Writing tests as coming from a rigorous, research-based standard setting method for alternate assessments. The results are reliable and consistent with those from similar standard setting workshops conducted by Pearson.

This report is suitable for use by ISBE staff and their technical experts for understanding and interpreting the IAA Standard Setting process and outcomes. Pearson will provide results from the report in alternate formats for use by ISBE in communicating the IAA standard setting process and outcomes as directed.

Chapter VIII

Spring 2008 Item Field Test

Field Test Design

Starting with the Spring 2007 IAA Writing operational administration in grades 5, 8, and 11, an embedded field test design has been used. Each operational IAA form includes five to ten scored items and one or two embedded field test items depending on subject. Seven forms are administered per subject per grade ($n = 200$ per form) with unique field test items in each form.

Forms Assignment

In order to use an equivalent groups design for item calibration and pre-equating, it was necessary to assign forms using a stratified random sample approach at the school level. The pre-id and ACC files for the Spring 2008 administration were used as the frame for form assignment. Spring 2008 field test form assignment followed the following steps:

1. Students were grouped by grade level and a separate file was created for each grade.
2. Within each grade file, students were aggregated by RCDTS_HOME.
3. For each RCDTS_HOME code, “total number of students,” “total number of white students,” and “percent non-white students” variables were created.
4. A forms spiraling file for each grade that included RCDTS_HOME code, “total number of students,” “percent non-white students” and “total number of forms to be assigned” was created and sorted prior to assigning forms.
5. The file was sorted in descending order by “percent non-white students.”
6. Next, the file was sorted in descending order by “total number of students” within “percent non-white students.”
7. Finally, the file was sorted by ascending RCDTS_HOME code within “total number of students.”
8. Forms were assigned to each RCDTS_HOME code in a spiraling design.

Analysis

Following the processing of answer documents, student demographic and item response data were transmitted to Pearson’s psychometric services division. Pearson psychometric staff had primary responsibility for analyzing IAA field test data to ensure accuracy and validity of scoring. Most of the psychometric work was carried out using SAS Version 9.1 and WINSTEPS Version 3.6, commercially available statistical analysis software. Traditional item analysis and data file QC analyses were conducted with SAS programs. Item response theory (IRT) analyses were conducted with the WINSTEPS program (Linacre, 2006). WINSTEPS allows for estimation of IRT item parameters for dichotomously or polytomously scored items. It has been thoroughly tested and is currently utilized by several high-stakes testing programs administered by Pearson.

All technical support and analyses were carried out in accordance with both the Standards for Educational and Psychological Testing, issued jointly by the American Educational Research Association, the American Psychological Association, the National Council on Measurement in Education, and the Pearson Quality Assurance Program. Pearson staff verified the IAA data and analysis process at several steps in the procedure. This included verification of the SAS and WINSTEPS programs prior to use on actual data through review by a second member of the psychometric services staff. Additionally, the output from the traditional and IRT item analysis programs were verified for out of range values and for consistent results across programs. Finally, the IRT calibrations were rerun independently by a second Pearson psychometrician.

Pearson conducted extensive statistical analyses on all pilot items. These analyses showed which items were at an appropriate difficulty level for the testing population and screened for differential item difficulty for subgroups in the student population. The analysis of the test data can be broken down into several components: 1) classical item analyses; 2) differential item functioning (DIF) analyses; and 4) calibration of items for bank values to be used in test construction. In the following sections, the analysis procedures for each component are described in detail. The results of these analyses are reported in Appendix H.

Classical Item Analyses

Classical item analyses involve computing, for every item on each form, a set of statistics based on classical test theory. Analyses were reported both overall and by various ethnic and gender groups as sample size permits. The classical item statistics that were calculated for the pilot test items within each form include:

- Number of students tested for each item, overall and by subgroup.
- Item Means. The mean raw item score was computed for each polytomous item and is analogous to the p-value for dichotomously scored items. This is a measure of the difficulty of an item, in classical test theory, and is indicated by the average raw score for an item across all students from the rubric ratings. For polytomously scored items, this statistic indicates the average rating earned on the item. Desired values generally fall within the range of: (minimum score + 2 standard errors of measure) and (maximum score – 2 standard errors of measure).
- Item Score Point Frequency Distributions. These data provide information about the effectiveness of the scoring rubric. The criterion used for judging this aspect of item functioning is that no rubric point would be used less than 5% of the time for an item in the total test population. Rubric point use by item was negatively skewed in every case.
- Item-to-Total Score Correlations. This statistic helps evaluate how well a item discriminates between high-performing and low-performing examinees. It is sometimes referred to as a discrimination index because it is an indicator of the degree to which students who do well on this content area also do well on this item. Items with negative or extremely low correlations (< 0.05) can indicate serious problems with the item itself or can indicate that students have not been taught the content. Due to the small number (6) and similarity of items, IAA item-total correlations tend to be higher than seen on longer tests with more heterogeneous items. Based on the range of polyserials produced

in field test analyses, and by psychometric convention, an indicator of poor discrimination was set to $r = 0.30$.

Results of each of these analyses are presented in Tables H.1 – H.21.

Differential Item Functioning (DIF) Analyses

DIF analyses for the Spring 2008 field test items were conducted using a multiple method approach as was done with the Spring 2007 field test data. DIF statistics were estimated for all major subgroups with sufficient sample size: Black, Hispanic, and Female. Items with statistically significant differences in performance were flagged so that items could be carefully examined for possible biased or unfair content that was undetected in earlier fairness and bias content review meetings held prior to form construction. DIF flags and item means by gender and ethnicity are reported for field test items by grade and subject in Tables H.22 – H.42.

IRT Analysis

- Item Fit Estimates. The extent to which the Rasch-Partial Credit model conformed to the data was estimated item by item. This diagnostic information was helpful in the selection of equating items eligible for psychometric equating.
 - INFIT MNSQ is the average of the INFIT mean-squares associated with the responses in each category. The expected values for all categories are 1.0.
- Category step values. These values are on the theta scale and represent the point on this scale at which one is more likely to receive the higher of two score point (e.g., the point on the ability continuum at which one is more likely to receive a score of 2 than a score of 1). This parameter indicates how difficult it is to observe a category, not how difficult it is to perform it.
- Average measure. The "average measure" for a category is the average ability of the people who respond in that category or to that distracter. $\text{Average measure} = \frac{\sum (B_n - D_i)}{\text{count of observations in category}}$. This is an empirical value. It is not a Rasch-model parameter.

From the IRT analysis, Pearson found that based on the average measure and category step values that the categories were properly ordered for all field test items and that the information provided across categories for all items appropriately covered the -3 to +3 range of the theta scale. In Table H.43 – H.63 the IRT results are presented for each field test item by grade and subject.

Data Review

All items field tested in Spring 2008 will go through data review in the Fall of 2008. This data review will conform to the process outlined in Chapter IV of this Technical Manual.

Item Bank

The items field tested in Spring 2008 that survive the data review process will be added to the IAA item bank for potential operational use in future test cycles.

Chapter IX

Validation Study

Introduction

According to science (Cronbach & Meehl, 1955), practice (Messick, 1989), and law (NCLB, 2002), reliability and validity are elements essential to defensible score interpretation and use for any test, including alternate assessments such as the Illinois Alternate Assessment (IAA).

Although an alternate assessment, compared to general assessments, presents unique challenges to establishing reliability and validity, it is, appropriately, held to no lower professional standards. Without adequate reliability and validity, there can be no assurance that the IAA is measuring student abilities in the same way across assessment items or across students, or that the IAA is truly measuring the extended standards for reading or math, and not variables unrelated to the standards and to subsequent academic performance.

Performance based measurement

Accurate estimation of the reliability and validity for any assessment relies on appropriate understanding, definition, and measurement of the construct of interest, or as posited by Dawis (1987), an existing, accurate *theory of the scale* for the assessment. In the case of the IAA, the theory of the scale was proposed a priori, established through content standard specification/alignment (i.e., extension of general education content standards); operationalized through item and test design, review, and analysis; and is the basis for estimation of the reliability and validity of the IAA. As is described in Chapters I-II of this report, the IAA assessment design focuses on measuring actual student performance elicited by a trained teacher on a specified set of content-valid IAA items using materials appropriate to the student's usual communication methods using a standardized scoring rubric. These factors are included in the IAA theory of the scale and are addressed in the validation design.

Measurement of actual performance (e.g., authentic assessment) is the gold standard of applied human behavior assessment and is the preferred method for decision making in medical, clinical and workplace settings where stakes are high, performance variance across individuals may be unknown, and high validity is required (Rosenthal & Rosnow, 1991). The keys to measurement of actual performance are: a) identifying the performance of interest to measure, b) understanding the performance of interest within a larger model of behavior and influencing factors, c) specifying an appropriate measurement model, and c) designing data collection that will best meet model requirements. Many models of human performance exist, from molecular cognitive models to molar models of human performance within organizations (e.g., Naylor, Pritchard, & Ilgen, 1984); selection of an appropriate model depends largely on the level of performance to be measured. For example, student performance related to demonstration of IAA content standard, grade-level knowledge is not at the molecular cognitive process level, or at the person interacting within the classroom level, but at the level of individual observable performance in response to IAA items. Because of the large variance in individual needs across students coming into the assessment situation for the IAA population, the most valid performance model is one that provides both the right type and right amount of standardization in the face of a plethora of meaningful individual differences dimensions. The most valid assessment of a common construct across students who are each unique in how they retrieve,

process and convey relevant information is to assess each on the construct using the modality that is appropriate for that student; construct relevant factors are held constant, or standardized, and construct irrelevant factors are allowed to vary according to the student needs. Based on our work with various relevant performance models, the basic structure of the IAA performance model was posited (Figure 1) as a guide for the validation study design. In this model, content validity and standardization is built into the IAA performance items, teacher training, administration materials, scoring rubric, and protocol. Flexibility is provided through each teacher's best judgment of a student's unique needs regarding assessment modality (i.e., mode of communication). Students interact with and respond to IAA performance items given their needs and through a knowledgeable teacher's administration. Teacher scoring is standardized through training to a protocol and use of a rubric validated through expert judgment and field testing. This is the basic framework of the IAA student performance model: Actual student performance is elicited for any given student in response to content valid IAA items administered in such a way that the most valid assessment of the given student's content knowledge is observed and scored in a standardized manner.

Also included in Figure 1 are the validation components for the performance model that provide the basis for the validity study described in this chapter of the Technical Manual. Two sets of specially trained subject matter experts (SMEs; second scorers and expert scorers) with sufficient knowledge of the IAA content, administration, and student population to be described as validation experts observe the totality of an IAA assessment item. The first set, second scorers, consisted of educators located within students' districts (typically district coordinators). These SMEs provide a second score (i.e., Student Score 2) for students' performance using the same materials and protocol as the teacher giving the first and primary score (i.e., Student Score 1) for the student assessment. The purpose of these scores was to establish the inter-rater reliability of the IAA. The second set of SMEs, expert scorers, consisted of four specifically selected individuals who met a set of pre-determined criteria that defined them as experts in the evaluation of the IAA testing population. These SMEs provide an additional score (i.e., Expert Score) for students' performance using the same materials and protocol as the teacher giving the first and primary score (i.e., Student Score 1) for the student assessment. They also documented the assessment situation, the assessment modality, and other pertinent information regarding the assessment item on a standardized validation survey. The correlation between Student Score 1 and Expert Score is presented in the Figure 1 as a validity coefficient "xy". This validation approach is based on the premise that a score given to a student performance by a trained, objective SME is a true performance score that may be used as an external criterion for estimating concurrent criterion validity, if the SME has observed the same student performance as the teacher providing the score to be validated. Support for this approach is provided through existing validation research in education and industry (see Suen, 1990). A unique feature of the Spring 2008 validation study is that each student in the validation sample was scored by two expert scores, a second scorer, and the primary scorer (teacher). This design allowed for the examination of the inter-rater reliability of Expert Scores and the relationship between Expert Scores and Student Score 2 in addition to computing the validity coefficient "xy".

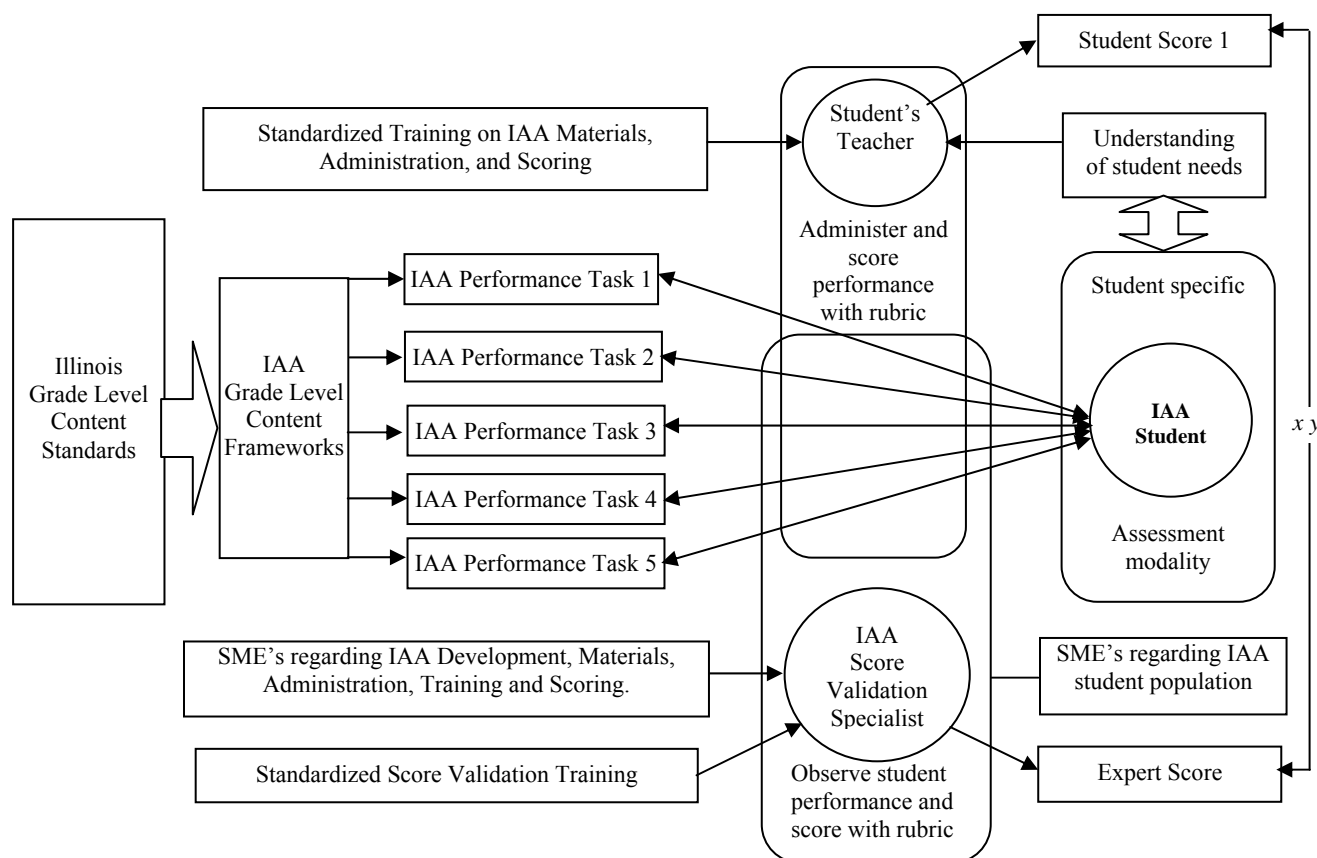


Figure 1. IAA performance model with validation component

IAA Reliability

Reliability of the IAA is a necessary precondition for validity. Reliability was built into the IAA in a formative manner, with evaluation provided in the thorough documentation of test material development, administration design, scoring process, and training efforts. As described in Chapters II and III of this report, IAA materials were designed with the goal of high inter- and intra-rater reliability in operational test administration use.

Reliability of the IAA was evaluated in a summative manner through analyses of data from the operational administration and the primary scorer/secondary scorer component of the validation study.

Reliability of the IAA tests in an actual operational setting--given student score variance and error variance associated with the 2008 operational administration and scoring, are presented as classical and conditional estimates of IAA test form reliability in an earlier Chapter of this Technical Manual (i.e., Chapter V). Coefficient alpha reliability estimates for the operational tests in each grade meet conventional guidelines for applied test reliability (i.e., $\alpha > .80$).

A total of 204 students received a Student Score 2 from a local second scorer. In order to establish the inter-rater reliability of the IAA tests Student Score 1 from the primary scorer and Student Score 2 from the second scorers were correlated. The results of these analyses suggest a

high degree of inter-rater reliability. Across grades correlations between Student Score 1 and Student Score 2 exceeded .80 for all subjects, and were approaching unity for most. The overall results by subject are summarized in Table 9.1; results by grade for mathematics, reading, science, and writing are shown in Tables 9.2 – 9.5 respectively.

Table 9.1. Spring 2008 IAA Inter-Rater Reliability by Subject

Overall		
Subject	Sample Size	Correlation
Math	184	0.978
Reading	186	0.975
Science	88	0.886
Writing	96	0.973

Table 9.2. Spring 2008 IAA Mathematics Inter-Rater Reliability by Grade

Mathematics		
Grade	Sample Size	Correlation
3	27	0.965
4	34	0.990
5	20	0.990
6	23	0.939
7	25	0.993
8	31	0.990
11	24	0.987

Table 9.3. Spring 2008 IAA Reading Inter-Rater Reliability by Grade

Reading		
Grade	Sample Size	Correlation
3	31	0.976
4	31	0.995
5	20	0.998
6	22	0.949
7	26	0.995
8	31	0.997
11	25	0.965

Table 9.4. Spring 2008 IAA Science Inter-Rater Reliability by Grade

Science		
Grade	Sample Size	Correlation
4	33	0.975
7	30	0.726
11	25	0.996

Table 9.5. Spring 2008 IAA Writing Inter-Rater Reliability by Grade

Writing		
Grade	Sample Size	Correlation
5	20	0.986
6	19	0.896
8	30	0.995

11	27	0.974
----	----	-------

IAA Validity

As implied by the IAA performance model in Figure 1 and posited by Messick (1989), validity of the assessment is built up through relevant, integrated factors. The validity of the IAA rests on the content frameworks, assessment materials, teacher training, scoring materials, appropriate flexibility of the assessment item given student needs, and the accuracy of teacher scoring. Throughout this Technical Manual, the validity of these various IAA tests has been presented through logical development processes and qualitative judgments. In this section we present another form of validation evidence: criterion-related validity. Based on reliable collection and scoring of IAA data, we propose that the relationship (i.e., xy) between Student Score 1 and Expert Score, as delineated in Figure 1, provides an estimate of the criterion-related validity of the IAA.

Methods

This study compared teacher ratings of student IAA performance to expert ratings of the same. As a first step, criteria for expert raters were defined:

1. Certified educator
2. Familiar with student population
3. Subject matter expert regarding IAA test design
4. Subject matter expert regarding IAA rubric
5. Agree to participate in larger research efforts
6. Ability to travel to schools during IAA testing window
7. Proof of insurance
8. Adhere to terms of confidentiality

The sampling plan was developed with the goals of providing adequate numbers of Expert Scores from a representative sample of IAA students to provide sufficient power to generalize results to the larger IAA population, while keeping within logistical and resource constraints for the study. With these goals in mind, the sampling plan included four expert scorers each providing Expert Scores on IAA Mathematics, Reading, and Science tests administered to each of 24 students. This plan provides 72 Expert Scores at the total test form level.

ISBE solicited nominations and selected from that group four SMEs who best met the criteria stated above from across the State. Pearson developed a sampling frame of schools from which to solicit participation. ISBE then recruited schools from the representative, purposeful sample developed by Pearson. The sample was based on demographic diversity of students, grade level diversity within school, strength of technology infrastructure, and proximity to SMEs.

A training program was developed by Pearson to prepare the SMEs to be consistent in their approach and scoring for the expert scoring task. In preparation for the training, SMEs were asked to review the IAA Implementation Manual, scoring rubric, score sheet, IAA sample items,

and the Online User's Guide at ISBE's IAA website. Group training for the four SMEs, conducted by Pearson and ISBE via teleconference, included review and group discussion of the test materials, test administration, and scoring process. Additionally, frame-of-reference training was provided to the group. Data collection methods were outlined and the SMEs reviewed the standardized survey for collecting information regarding the assessment event for each student.

Expert Scores were collected during the Spring 2008 IAA operational test window. Coordination of data collection activities among teachers, SMEs, and participating schools was a joint effort between ISBE, the SMEs, and Pearson. These Expert Scores were merged with operational test scores for students in the sample. Analyses of the merged data were conducted as described in the Results section of this Chapter.

Results

The sample characteristics for the validation study are presented by grade in Table 9.6, with comparison percentages from the Spring 2008 total IAA student test population.

Table 9.6. Spring 2008 IAA student population and validation sample characteristics

Spring 08 IAA Population			Validation Sample		
N	M	F	N	M	F
10015	63%	37%	22	64%	36%

Expert Score Inter-Rater Reliability

As mentioned previously, reliability is a necessary pre-condition for validity. Based on the reliability evidence presented in Chapter V of this Technical Manual and the inter-rater reliability evidence presented earlier in this Chapter, one can conclude that primary IAA test scores show strong reliability. Before examining evidence of criterion-related validity based on Expert Scores, it is important to show that these Expert Scores are sufficiently reliable to serve as a criterion variable. For this reason each student was observed by two expert scorers. The correlations between expert scorers indicate a high degree of reliability in their scoring. These results are summarized in table 9.7.

Table 9.7. Inter-Rater Reliability of Spring 2008 Expert Scores

Subject	Sample Size	Correlation
Overall	59	0.998
Mathematics	19	0.989
Reading	20	0.996
Science	20	0.999

Correlations with Expert Scores

The correlation between Student Score 1 and Expert Score was computed for across IAA subjects and for each subject. Given the nearly identical ratings provided by expert scorers the two scores provided for each student were averaged to create the Expert Score criterion variable. These correlations show a strong positive relationship between the sets of scores both across subjects and by subject. In Table 9.8 correlations between primary and expert scorer ratings are presented.

Table 9.8. Correlation between Student Score 1 and Expert Score for all Spring 08 IAA Tests

Subject	Sample Size	Correlation
Overall	59	0.944
Mathematics	19	0.986
Reading	20	0.990
Science	20	0.781

The correlation between Student Score 2 and Expert Score was calculated. The second scorers, like expert scores, were provided with specific training, and as such these results provide context for the relationships between primary and expert scores. The results of these analyses showed nearly perfect correspondence between second scorers and expert scorers. These results are summarized in Table 9.9.

Table 9.9. Correlation between Student Score 2 and Expert Score for all Spring 08 IAA Tests

Subject	Sample Size	Correlation
Overall	59	0.997
Mathematics	19	0.997
Reading	20	0.988
Science	20	0.999

Discussion

The validity evidence from the study is very clear: The teacher scores on the IAA tests are valid. The validity results found in this study are as strong as any in published validation research. The validity coefficients based on the correlation between Student Score 1 and Expert Score range from 0.781 to 0.990 by subject and the correlation is 0.944 across subjects. The correlations between Student Score 2 and Expert Score suggest that primary scorers (teachers) may still benefit from additional training – especially with respect to Science. The correlation between expert scorers and primary scorers is 0.781 for Science, but the correlation between experts and second scorers for this subject is 0.999. The specialized training provided to second scorers may have played some role in the greater correspondence between their scores and those of the expert scorers. Nonetheless, both the inter-rater reliability and criterion-related validity results suggest that the IAA provides accurate assessment of the performance of students in the 1% population.

This study faces three major limitations. The design of this study necessitated that a total of three scorers in addition to the primary scorer be present to observe students' performance on the IAA. Given the sensitivity of students in this population to changes in environment it is possible that the addition of these individuals had an affect on student performance (e.g., caused a distraction). After all ratings were completed the four expert scorers participated in a focus group led by representatives from Pearson and ISBE. The purpose of this focus group was to determine what went well and what could be improved with regard to the execution of this study. The issue of rater presence affecting student performance was specifically addressed during this meeting. The expert raters saw no reason to believe that their presence has an adverse affect on student performance. Additionally, during training, expert and second scorers were instructed to excuse themselves from the testing environment if they felt their presence was causing a disturbance. Given the presence of all four scorers in such close proximity it was also possible that the presence of other scorers would cause rater bias. The expert raters reported that this was also not an issue during the study. The final limitation concerns the IAA score distributions. The very skewed score distributions for all IAA tests inflated the agreement results due to a score

point of “4” always being most probable regardless of rater. This inflation explains why the validity results reported in this Chapter appear so much more favorable than those typically seen in criterion-related validation research.

References:

- American Educational Research Association (AERA), American Psychological Association (APA), and the National Council on Measurement in Education (NCME) (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.
- Cronbach, L. J. & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-302.
- Dawis, R. (1987). A theory of work adjustment. In B. Bolton (Ed.). *Handbook on the measurement and evaluation in rehabilitation* (2nd ed.) (pp. 207-217). Baltimore: Paul H. Brooks.
- Denzin, N. & Lincoln, Y. (2007). *Collecting and Interpreting Qualitative Materials* (3rd ed.). Thousand Oaks, CA: Sage.
- Kingston, N. M., Kahl, S. R., Sweeney, K. P., & Bay, L. (2001). Setting Performance Standards Using the Body of Work Method. In G.J. Cizek (Ed.), *Setting Performance Standards: Concepts, Methods, and Perspectives*, (pp. 219-248). Lawrence Erlbaum Associates, Mahwah, NJ.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating: Methods and practices* (2nd ed). New York: Springer-Verlag.
- Linacre, J. M. (2006). *WINSTEPS: Rasch measurement, Version 3.6* [Computer Software]. Chicago, IL: Winsteps.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 13-103). New York: Macmillan.
- Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives, pp. 249 – 282. In G.J. Cizek (Ed.) *Setting Performance Standards: Concepts, Methods, and Perspectives*, Mahwah, NJ: Lawrence Erlbaum & Associates.
- Naylor, J. C., & Ilgen, D. R. (1984). Goal setting: A theoretical analysis of a motivational technology. In L. L. Cummings & B. M. Staw (Eds.), *Research in Organizational Behavior* (Vol. 6). Greenwich, CT: JAI Press.
- No Child Left Behind Act of 2001, 20 U.S.C. ? 6301 et seq (2001) (PL 107-110).

- Rosenthal, R. & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw Hill.
- Rosnow, R. L., & Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people's published data: General procedures for research consumers. *Psychological Methods*, 1, 331-340.
- Schafer, William D. (2005). Technical documentation for alternate assessments. *Practical Assessment Research & Evaluation*, 10(10). [Available online: <http://pareonline.net/getvn.asp?v=10&n=10>].
- Suen, H. K. (1990). *Principles of test theories*. Hillsdale, NJ: Erlbaum.

APPENDICES TO
Illinois Alternate Assessment (IAA)
2007-2008
Annual Technical Manual

Appendix A

Sample IAA Task

Grade 6

Assessment Objective: Solve problems and number sentences involving addition, subtraction, multiplication, and division using whole numbers.

Mathematics Sample Task 02

Teacher Instruction:

Provide the student with a group of 17 tiles and another group of 6 tiles. Provide the student with the addition problem $17 + 6$. Provide the student with three answer choices: 13, 23, and 25. Ask the student, "How many tiles are there all together?" (Correct Answer: 23)

Student Task:

Solve the addition problem $17 + 6$.

Materials:

**Materials typically used for addition problems
23 tiles
The problem $17 + 6$
Three answer choices: 13, 23, and 25**

Appendix B

IAA Scoring Rubric

IAA PERFORMANCE-BASED TASK SCORING RUBRIC

<i>Level 4:</i>	<i>Level 3:</i>	<i>Level 2:</i>	<i>Level 1:</i>
<p>The student correctly performs the task without assistance or with a single repetition of instructions or refocusing.</p> <ul style="list-style-type: none"> The student responds correctly to the task when presented as it is written in the instructions with the necessary materials. If the student does not respond independently or responds incorrectly to the initial presentation of the task when given adequate wait time, the teacher repeats the instructions and/or refocuses the student's attention. <p><i>The student then responds correctly.</i></p>	<p>The student correctly performs the task with general prompts.</p> <ul style="list-style-type: none"> If the student responds incorrectly to the task at Level 4 when given adequate wait time, the teacher provides additional information or adds prompts about the expected response from the student such as: <ul style="list-style-type: none"> Elaborating or providing additional clarifying information on directions or expected response. Demonstrating a like response such as, "This is a picture of a dog. Show me a picture of a cat." Providing examples but not modeling the correct response. <p><i>The student then responds correctly.</i></p>	<p>The student correctly performs the task with specific prompts.</p> <ul style="list-style-type: none"> If the student responds incorrectly to the task at Level 3 when given adequate wait time, the teacher provides specific prompts to direct the student's correct response such as: <ul style="list-style-type: none"> Modeling exact response, "This is a picture of a dog, what is this?" (Show a picture of a dog). After physically guiding the student to the correct response such as using hand over hand, the student then indicates the correct answer in his/her mode of communication. <p><i>The student responds correctly after being given the correct answer.</i></p>	<p>The student does not perform the task at Level 2 or provides an incorrect response despite Level 2 support.</p> <p><i>The student does not respond or does not respond correctly. Teacher demonstrates response and moves on to the next task.</i></p>

Illinois State Board of Education has adapted this rubric from the Colorado Student Assessment Program Alternate Level of Independence Performance Rubric. ISBE August 31, 2006

Appendix C
IAA Paper Scoring Sheet

DO NOT RETURN THIS SCORE SHEET TO PEARSON FOR PROCESSING.All scores must be submitted through Pearson's online system by 11:59 pm on April 11th, 2008.**IAA 2008 – SCORE SHEET**

Teacher Name (Last, First): _____

STUDENT DEMOGRAPHICS

Name (Last, First): _____

Grade Level (circle one): 3 4 5 6 7 8 11

Gender (circle one): Male or Female

Date of Birth (mm/dd/yyyy): _____ / _____ / _____

SIS ID Number: _____

Ethnicity (circle one): American Indian Asian/Pacific Islander Black/African American

White/Caucasian

Hispanic

Multiracial/Ethnic

Home School RCDTS Code: _____

Testing School RCDTS Code: _____

Test Form Number (circle one): 1 2 3 4 5 6 7

Reading											
Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10	Task 11	
Score											

Mathematics												
Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9	Task 10	Task 11	Task 12	
Score												

Science								
Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	Task 7	Task 8	
Score								

Writing						
Task 1	Task 2	Task 3	Task 4	Task 5	Task 6	
Score						

Appendix D
Item Analysis – Spring 2008 Operational Forms

Table D.1. Percentage of Students Earning Each Score Point - Grade 3 Mathematics					
		Score Point			
Item	N	1	2	3	4
3M002	1250	8	5	5	81
3M013	1250	14	14	17	54
3M022	1250	19	26	21	34
3M029	1250	18	26	26	31
3M036	1250	13	11	11	64
3M038	1250	11	12	17	60
3M047	1250	16	18	17	49
3M049	1250	9	8	9	74
3M056	1250	17	18	23	42
3M061	1249	15	18	20	46

Table D.2. Percentage of Students Earning Each Score Point - Grade 4 Mathematics					
		Score Point			
Item	N	1	2	3	4
4M002	1381	8	13	11	68
4M005	1381	9	10	10	70
4M024	1381	16	26	20	38
4M031	1381	12	14	13	61
4M036	1381	9	12	11	69
4M040	1381	9	14	17	60
4M045	1381	15	25	24	36
4M049	1381	9	11	19	61
4M056	1381	9	7	10	73
4M061	1381	12	13	15	59

Table D.3. Percentage of Students Earning Each Score Point - Grade 5 Mathematics					
		Score Point			
Item	N	1	2	3	4
5M002	1340	12	21	23	44
5M012	1340	19	29	25	27
5M016	1340	17	21	22	39
5M026	1340	9	17	19	55
5M035	1340	8	11	10	71
5M039	1340	9	11	13	66
5M043	1340	16	27	22	34
5M051	1340	8	13	14	64
5M056	1340	10	12	14	64
5M062	1340	10	16	20	54

Table D.4. Percentage of Students Earning Each Score Point - Grade 6 Mathematics					
		Score Point			
Item	N	1	2	3	4
6M005	1502	10	17	20	53
6M015	1502	14	25	25	36
6M021	1502	11	17	22	50
6M026	1502	8	9	11	72
6M028	1502	10	12	13	65
6M035	1502	11	14	17	59
6M043	1502	9	12	16	64
6M053	1502	7	9	12	73
6M061	1502	10	14	16	61
6M064	1502	9	10	10	71

Table D.5. Percentage of Students Earning Each Score Point - Grade 7 Mathematics					
		Score Point			
Item	N	1	2	3	4
7M003	1491	12	17	22	49
7M012	1491	16	32	26	26
7M023	1491	13	16	20	51
7M024	1491	10	17	16	57
7M030	1491	13	19	20	48
7M039	1491	19	36	23	21
7M043	1490	14	20	22	43
7M054	1490	8	10	10	72
7M057	1490	15	25	21	38
7M066	1490	17	20	20	43

Table D.6. Percentage of Students Earning Each Score Point - Grade 8 Mathematics					
		Score Point			
Item	N	1	2	3	4
8M001	1530	8	13	14	65
8M010	1530	13	21	17	49
8M021	1530	7	11	13	69
8M025	1530	7	9	11	73
8M029	1530	12	20	23	45
8M037	1530	12	19	19	49
8M045	1530	14	24	28	34
8M049	1530	8	9	11	72
8M053	1530	10	14	13	63
8M058	1530	12	17	18	53

Table D.7. Percentage of Students Earning Each Score Point - Grade 11 Mathematics					
		Score Point			
Item	N	1	2	3	4
11M002	1149	11	14	20	55
11M017	1149	13	18	28	41
11M024	1149	16	28	29	28
11M027	1149	14	25	25	36
11M033	1149	15	20	24	41
11M037	1149	11	14	20	55
11M042	1149	11	15	21	53
11M049	1148	8	9	16	67
11M059	1148	11	10	17	62
11M061	1148	9	10	16	64

Table D.8. Percentage of Students Earning Each Score Point - Grade 3 Reading					
		Score Point			
Item	N	1	2	3	4
3R001	1253	11	13	20	56
3R008	1253	9	9	16	66
3R011	1253	10	8	12	70
3R018	1253	12	13	16	59
3R022	1253	11	14	18	57
3R029	1253	10	12	19	59
3R044	1253	10	10	13	68
3R046	1253	13	12	20	56
3R051	1253	18	20	17	45

Table D.9. Percentage of Students Earning Each Score Point - Grade 4 Reading					
		Score Point			
Item	N	1	2	3	4
4R004	1381	8	11	12	69
4R008	1381	7	10	15	68
4R011	1381	8	11	12	69
4R018	1381	10	11	17	61
4R022	1381	11	13	17	59
4R035	1381	10	15	20	55
4R040	1381	11	17	19	53
4R049	1381	9	9	13	69
4R054	1381	17	25	24	33

Table D.10. Percentage of Students Earning Each Score Point - Grade 5 Reading					
		Score Point			
Item	N	1	2	3	4
5R002	1339	9	11	15	65
5R012	1339	8	11	17	63
5R013	1339	10	14	18	58
5R022	1339	11	16	21	52
5R025	1339	7	9	13	71
5R032	1339	11	18	24	48
5R043	1339	10	12	18	60
5R056	1339	9	11	19	60
5R062	1339	14	21	21	45

Table D.11. Percentage of Students Earning Each Score Point - Grade 6 Reading					
		Score Point			
Item	N	1	2	3	4
6R003	1501	7	10	17	66
6R008	1501	9	13	19	59
6R015	1501	13	15	17	55
6R021	1501	10	16	24	50
6R037	1501	13	22	26	39
6R040	1501	11	16	20	52
6R047	1501	8	11	17	64
6R050	1500	8	10	14	68
6R063	1500	12	18	17	53

Table D.12. Percentage of Students Earning Each Score Point - Grade 7 Reading					
		Score Point			
Item	N	1	2	3	4
7R002	1493	8	11	17	64
7R005	1493	9	12	15	63
7R011	1493	10	13	17	61
7R021	1493	7	12	14	66
7R029	1493	11	13	16	60
7R043	1493	13	19	17	51
7R046	1492	9	13	16	62
7R054	1492	12	15	19	54
7R057	1492	16	21	23	41

Table D.13. Percentage of Students Earning Each Score Point - Grade 8 Reading					
		Score Point			
Item	N	1	2	3	4
8R001	1531	5	9	10	76
8R008	1531	8	12	17	64
8R019	1530	12	20	23	46
8R022	1531	10	17	23	50
8R032	1531	7	9	11	73
8R039	1531	7	14	19	59
8R044	1530	9	14	17	60
8R046	1530	10	14	18	59
8R051	1530	9	13	18	59

Table D.14. Percentage of Students Earning Each Score Point - Grade 11 Reading					
		Score Point			
Item	N	1	2	3	4
11R003	1151	5	6	9	79
11R007	1151	8	9	16	67
11R014	1151	7	7	11	75
11R020	1151	6	6	11	77
11R024	1151	7	7	11	75
11R035	1151	8	10	12	70
11R038	1151	9	10	19	63
11R044	1151	9	10	15	66
11R051	1151	7	9	15	69

Table D.15. Percentage of Students Earning Each Score Point - Grade 4 Science					
		Score Point			
Item	N	1	2	3	4
4S008	1377	13	17	17	54
4S014	1377	9	12	12	67
4S026	1377	13	17	14	56
4S035	1377	8	12	13	67
4S042	1377	10	12	10	68
4S057	1377	10	12	13	65

Table D.16. Percentage of Students Earning Each Score Point - Grade 7 Science					
		Score Point			
Item	N	1	2	3	4
7S003	1488	8	11	14	68
7S015	1488	5	11	12	72
7S024	1488	11	13	14	62
7S037	1488	12	18	21	49
7S045	1488	9	13	14	64
7S052	1488	11	15	17	58

Table D.17. Percentage of Students Earning Each Score Point - Grade 11 Science					
		Score Point			
Item	N	1	2	3	4
11S008	1147	11	10	20	60
11S016	1147	11	15	25	49
11S023	1147	8	8	14	70
11S035	1147	14	16	19	51
11S042	1147	10	12	20	57
11S045	1147	6	5	10	79

Table D.18. Percentage of Students Earning Each Score Point - Grade 5 Writing					
		Score Point			
Item	N	1	2	3	4
5W021	1338	9	10	13	68
5W029	1338	9	19	19	54
5W100	1338	13	18	15	54
5W123	1338	12	17	22	49
5W130	1338	14	17	18	51

Table D.19. Percentage of Students Earning Each Score Point - Grade 6 Writing					
		Score Point			
Item	N	1	2	3	4
6W003	1501	15	17	21	47
6W012	1501	16	21	23	40
6W018	1501	17	17	21	46
6W021	1501	8	9	12	71
6W030	1501	18	29	24	29

Table D.20. Percentage of Students Earning Each Score Point - Grade 8 Writing					
		Score Point			
Item	N	1	2	3	4
8W038	1526	10	13	20	57
8W043	1529	10	13	16	60
8W052	1529	8	11	10	71
8W054	1529	11	15	16	57
8W060	1529	9	12	13	66

Table D.21. Percentage of Students Earning Each Score Point - Grade 11 Writing					
		Score Point			
Item	N	1	2	3	4
9W028	1150	8	12	18	62
9W101	1150	9	12	17	62
9W107	1150	8	9	16	67
9W110	1150	9	11	20	60
9W117	1150	9	9	15	67

Table D.22. Item Content, Item-Total Correlation, and Mean - Grade 3 Mathematics							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
3M002	1250	1	6	A	1	0.78	3.59
3M013	1250	2	6	A	7	0.82	3.11
3M022	1250	3	6	B,C	10	0.69	2.71
3M029	1250	5	7	A,B,C	1	0.72	2.70
3M036	1250	4	7	A,B,C	5	0.80	3.26
3M038	1250	6	8	A	1	0.78	3.26
3M047	1250	7	8	C,D	4	0.78	2.98
3M049	1250	8	9	A	1	0.77	3.49
3M056	1250	11	9	B	9	0.74	2.90
3M061	1249	12	10	A,B	1	0.78	2.97

Table D.23. Item Content, Item-Total Correlation, and Mean - Grade 4 Mathematics

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
4M002	1381	1	6	A	1	0.81	3.39
4M005	1381	2	6	A	2	0.81	3.41
4M024	1381	3	6	B,C	11	0.67	2.80
4M031	1381	4	7	A,B,C	2	0.80	3.23
4M036	1381	5	7	A,B,C	4	0.80	3.40
4M040	1381	6	8	A	1	0.76	3.29
4M045	1381	7	8	C,D	7	0.67	2.80
4M049	1381	8	9	A	1	0.74	3.33
4M056	1381	11	9	B	12	0.80	3.48
4M061	1381	12	10	A,B	1	0.83	3.21

Table D.24. Item Content, Item-Total Correlation, and Mean - Grade 5 Mathematics

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
5M002	1340	1	6	A	1	0.71	2.98
5M012	1340	2	6	A	10	0.68	2.61
5M016	1340	3	6	B,C	12	0.73	2.83

Table D.24. Item Content, Item-Total Correlation, and Mean - Grade 5 Mathematics

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
5M026	1340	4	7	A,B,C	2	0.71	3.20
5M035	1340	5	7	A,B,C	4	0.74	3.45
5M039	1340	6	8	A	1	0.78	3.36
5M043	1340	7	8	C,D	8	0.67	2.74
5M051	1340	8	9	A	1	0.76	3.34
5M056	1340	11	9	B	14	0.76	3.31
5M062	1340	12	10	A,B	1	0.74	3.18

Table D.25. Item Content, Item-Total Correlation, and Mean - Grade 6 Mathematics

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
6M005	1502	1	6	A	4	0.72	3.16
6M015	1502	2	6	B,C	12	0.70	2.83
6M021	1502	3	7	A,B,C	1	0.76	3.11
6M026	1502	4	7	A,B,C	3	0.78	3.47
6M028	1502	5	8	A	1	0.81	3.34
6M035	1502	6	8	B	6	0.83	3.24
6M043	1502	7	9	A	2	0.77	3.35

Table D.25. Item Content, Item-Total Correlation, and Mean - Grade 6 Mathematics

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
6M05 3	150 2	8	9	B	12	0.80	3.50
6M06 1	150 2	11	10	A,B	1	0.84	3.28
6M06 4	150 2	12	10	C	6	0.82	3.42

Table D.26. Item Content, Item-Total Correlation, and Mean - Grade 7 Mathematics

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
7M00 3	149 1	1	6	A	3	0.72	3.08
7M01 2	149 1	2	6	D	16	0.65	2.63
7M02 3	149 1	4	7	A,B,C	1	0.77	3.08
7M02 4	149 1	3	7	A,B,C	3	0.72	3.19
7M03 0	149 1	5	8	A	1	0.76	3.02
7M03 9	149 1	6	8	C,D	12	0.61	2.47
7M04 3	149 0	7	9	A	5	0.76	2.95
7M05 4	149 0	8	9	B	14	0.74	3.46
7M05 7	149 0	11	10	A,B	1	0.70	2.82
7M06 6	149 0	12	10	C	7	0.75	2.90

Table D.27. Item Content, Item-Total Correlation, and Mean - Grade 8 Mathematics

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
8M001	1530	1	6	A	3	0.74	3.35
8M010	1530	2	6	B,C	9	0.77	3.01
8M021	1530	3	7	A,B,C	1	0.78	3.44
8M025	1530	4	7	A,B,C	3	0.77	3.50
8M029	1530	5	8	A	1	0.72	3.00
8M037	1530	6	8	C,D	13	0.78	3.05
8M045	1530	7	9	A	5	0.69	2.81
8M049	1530	8	9	B	10	0.80	3.46
8M053	1530	11	10	A,B	1	0.81	3.29
8M058	1530	12	10	A,B	3	0.79	3.13

Table D.28. Item Content, Item-Total Correlation, and Mean - Grade 11 Mathematics

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
11M002	1149	1	6	A	1	0.76	3.18
11M017	1149	2	6	B,C	13	0.72	2.98
11M024	1149	3	6	D	18	0.66	2.69

Table D.28. Item Content, Item-Total Correlation, and Mean - Grade 11 Mathematics							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
11M02 7	114 9	4	7	A,B,C	1	0.71	2.82
11M03 3	114 9	5	7	A,B,C	3	0.75	2.90
11M03 7	114 9	6	8	A	4	0.76	3.18
11M04 2	114 9	7	8	B	12	0.76	3.16
11M04 9	114 8	8	9	A	6	0.70	3.43
11M05 9	114 8	11	9	A	7	0.75	3.31
11M06 1	114 8	12	10	A,B	1	0.76	3.35

Table D.29. Item Content, Item-Total Correlation, and Mean - Grade 3 Reading							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
3R001	1253	2	1	A	3	0.79	3.21
3R008	1253	3	1	A	8	0.84	3.38
3R011	1253	4	1	A	11	0.84	3.42
3R018	1253	5	1	B,C	13	0.82	3.22
3R022	1253	6	1	B,C	14	0.82	3.23
3R029	1253	7	1	C	20	0.81	3.27
3R044	1253	1	2	A	7	0.82	3.38
3R046	1253	11	2	A	7	0.79	3.19
3R051	1253	10	2	B	10	0.75	2.90

Table D.30. Item Content, Item-Total Correlation, and Mean - Grade 4 Reading							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
4R004	1381	1	1	A	4	0.80	3.42
4R008	1381	2	1	A	7	0.82	3.45
4R011	1381	3	1	B,C	9	0.82	3.42
4R018	1381	4	1	B,C	10	0.77	3.29

Table D.30. Item Content, Item-Total Correlation, and Mean - Grade 4 Reading							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
4R02 2	138 1	5	1	B,C	15	0.78	3.24
4R03 5	138 1	6	1	C	25	0.75	3.19
4R04 0	138 1	7	2	A	1	0.77	3.14
4R04 9	138 1	10	2	A	6	0.81	3.42
4R05 4	138 1	11	2	B	13	0.66	2.74

Table D.31. Item Content, Item-Total Correlation, and Mean - Grade 5 Reading

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
5R00 2	133 9	1	1	A	2	0.79	3.37
5R01 2	133 9	2	1	B,C	7	0.78	3.36
5R01 3	133 9	3	1	B,C	8	0.80	3.24
5R02 2	133 9	4	1	B,C	13	0.77	3.13
5R02 5	133 9	5	1	C	16	0.81	3.48
5R03 2	133 9	6	1	C	20	0.72	3.09
5R04 3	133 9	7	2	A	1	0.80	3.29
5R05 6	133 9	10	2	A	8	0.78	3.30
5R06 2	133 9	11	2	B	14	0.72	2.96

Table D.32. Item Content, Item-Total Correlation, and Mean - Grade 6 Reading

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
6R00 3	150 1	1	1	A	3	0.74	3.42
6R00 8	150 1	2	1	A	5	0.76	3.29
6R01 5	150 1	3	1	B	11	0.82	3.14
6R02 1	150 1	4	1	B	12	0.75	3.15

Table D.32. Item Content, Item-Total Correlation, and Mean - Grade 6 Reading							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
6R037	1501	5	1	C	20	0.66	2.92
6R040	1501	6	1	C	22	0.74	3.14
6R047	1501	7	2	A	1	0.78	3.37
6R050	1500	10	2	A	7	0.81	3.42
6R063	1500	11	2	A	14	0.75	3.11

Table D.33. Item Content, Item-Total Correlation, and Mean - Grade 7 Reading							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
7R002	1493	1	1	A	3	0.77	3.38
7R005	1493	2	1	A	5	0.79	3.33
7R011	1493	3	1	B	9	0.80	3.29
7R021	1493	4	1	C	15	0.76	3.39
7R029	1493	5	1	C	17	0.81	3.25
7R043	1493	6	1	C	22	0.77	3.06
7R046	1492	7	2	A	1	0.81	3.30
7R054	1492	10	2	A	6	0.80	3.15
7R057	1492	11	2	A	8	0.74	2.89

Table D.34. Item Content, Item-Total Correlation, and Mean - Grade 8 Reading							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
8R001	1531	1	1	A	3	0.78	3.57
8R008	1531	2	1	B	8	0.77	3.36
8R019	1530	3	1	B	10	0.77	3.03
8R022	1531	4	1	B	12	0.74	3.12

Table D.34. Item Content, Item-Total Correlation, and Mean - Grade 8 Reading							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
8R03 2	153 1	5	1	C	16	0.79	3.50
8R03 9	153 1	6	1	C	23	0.76	3.31
8R04 4	153 0	7	2	A	1	0.81	3.28
8R04 6	153 0	10	2	A	4	0.82	3.26
8R05 1	153 0	11	2	A	6	0.81	3.28

Table D.35. Item Content, Item-Total Correlation, and Mean - Grade 11 Reading							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
11R003	1151	1	1	A	1	0.83	3.62
11R007	1151	2	1	A	2	0.81	3.42
11R014	1151	3	1	C	8	0.83	3.56
11R020	1151	4	1	C	9	0.83	3.58
11R024	1151	5	1	C	13	0.86	3.53
11R035	1151	6	1	C	14	0.83	3.43
11R038	1151	7	1	C	16	0.76	3.36
11R044	1151	10	1	C	22	0.80	3.38
11R051	1151	11	1	C	25	0.80	3.46

Table D.36. Item Content, Item-Total Correlation, and Mean - Grade 4 Science							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
4S008	1377	2	11	B	5	0.79	3.11
4S014	1377	3	12	A	3	0.85	3.37
4S026	1377	7	12	C	17	0.83	3.13
4S035	1377	1	12	E	40	0.82	3.38

Table D.36. Item Content, Item-Total Correlation, and Mean - Grade 4 Science							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
4S04 2	137 7	8	12	F	47	0.82	3.37
4S05 7	137 7	4	13	B	13	0.80	3.33

Table D.37. Item Content, Item-Total Correlation, and Mean - Grade 7 Science							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
7S00 3	148 8	2	11	A	2	0.82	3.41
7S01 5	148 8	1	12	A	1	0.80	3.52
7S02 4	148 8	7	12	C	49	0.82	3.28
7S03 7	148 8	3	12	E	85	0.75	3.07
7S04 5	148 8	8	12	F	91	0.81	3.33
7S05 2	148 8	4	13	A	1	0.82	3.22

Table D.38. Item Content, Item-Total Correlation, and Mean - Grade 11 Science							
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
11S00 8	114 7	2	11	A	7	0.78	3.30
11S01 6	114 7	1	12	A	25	0.73	3.13
11S02 3	114 7	3	12	B	31	0.80	3.46

Table D.38. Item Content, Item-Total Correlation, and Mean - Grade 11 Science

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
11S035	1147	7	12	D	76	0.77	3.08
11S042	1147	8	12	F	**	0.77	3.25
11S045	1147	4	13	A	1	0.81	3.62

Table D.39. Item Content, Item-Total Correlation, and Mean - Grade 5 Writing

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
5W021	1338	1	3	B	15	0.81	3.39
5W029	1338	2	3	B	21	0.79	3.16
5W100	1338	6	3	A	9	0.80	3.11
5W123	1338	3	3	A	10	0.81	3.09
5W130	1338	5	3	A	1	0.81	3.05

Table D.40. Item Content, Item-Total Correlation, and Mean - Grade 6 Writing

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
6W003	1501	5	3	A	1	0.83	2.99
6W012	1501	3	3	A	9	0.78	2.87
6W018	1501	6	3	A	10	0.84	2.97

Table D.40. Item Content, Item-Total Correlation, and Mean - Grade 6 Writing

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
6W021	1501	1	3	B,C	15	0.75	3.47
6W030	1501	2	3	B,C	22	0.76	2.64

Table D.41. Item Content, Item-Total Correlation, and Mean - Grade 8 Writing

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
8W038	1526	6	3	B,C	38	0.81	3.24
8W043	1529	3	3	A	10	0.84	3.26
8W052	1529	1	3	A	1	0.85	3.44
8W054	1529	2	3	A	9	0.83	3.21
8W060	1529	5	3	B,C	22	0.85	3.36

Table D.42. Item Content, Item-Total Correlation, and Mean - Grade 11 Writing

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
9W028	1150	2	3	B	28	0.81	3.35
9W101	1150	3	3	A	9	0.83	3.32
9W107	1150	6	3	A	14	0.85	3.41

**Table D.42. Item Content, Item-Total Correlation, and Mean -
Grade 11 Writing**

Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean
9W11 0	115 0	1	3	A	21	0.80	3.31
9W11 7	115 0	5	3	B,C	35	0.83	3.41

**Table D.43. Item Means by Gender and Ethnicity –
Grade 3 Mathematics**

Item	Female		Male		White		Black		Hispanic	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
3M00 2	436	3.52	81 4	3.63	66 5	3.69	25 9	3.47	226	3.43
3M01 3	436	3.05	81 4	3.14	66 5	3.18	25 9	3.02	226	3.03
3M02 2	436	2.67	81 4	2.73	66 5	2.78	25 9	2.49	226	2.69
3M02 9	436	2.66	81 4	2.72	66 5	2.85	25 9	2.39	226	2.59
3M03 6	436	3.29	81 4	3.24	66 5	3.37	25 9	3.12	226	3.11
3M03 8	436	3.24	81 4	3.28	66 5	3.37	25 9	3.10	226	3.11
3M04 7	436	2.97	81 4	2.99	66 5	3.06	25 9	2.79	226	2.89
3M04 9	436	3.43	81 4	3.52	66 5	3.60	25 9	3.37	226	3.31
3M05 6	436	2.84	81 4	2.94	66 5	3.05	25 9	2.66	226	2.72
3M06 1	436	2.99	81 3	2.97	66 4	3.02	25 9	2.90	226	2.92

Table D.44. Item Means by Gender and Ethnicity – Grade 4 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
4M002	479	3.32	902	3.42	776	3.41	296	3.39	225	3.33
4M005	479	3.29	902	3.48	776	3.46	296	3.37	225	3.30
4M024	479	2.71	902	2.85	776	2.82	296	2.68	225	2.81
4M031	479	3.22	902	3.23	776	3.25	296	3.17	225	3.20
4M036	479	3.36	902	3.42	776	3.42	296	3.36	225	3.37
4M040	479	3.21	902	3.34	776	3.32	296	3.31	225	3.16
4M045	479	2.72	902	2.85	776	2.80	296	2.74	225	2.89
4M049	479	3.23	902	3.38	776	3.40	296	3.20	225	3.24
4M056	479	3.47	902	3.48	776	3.49	296	3.44	225	3.52
4M061	479	3.16	902	3.24	776	3.24	296	3.16	225	3.15

Table D.45. Item Means by Gender and Ethnicity – Grade 5 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
5M002	486	2.99	854	2.97	733	3.01	288	2.90	223	3.02
5M012	486	2.53	854	2.65	733	2.66	288	2.53	223	2.61
5M016	486	2.83	854	2.83	733	2.87	288	2.72	223	2.85

Table D.45. Item Means by Gender and Ethnicity – Grade 5 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
5M026	486	3.16	854	3.23	733	3.28	288	3.11	223	3.09
5M035	486	3.47	854	3.44	733	3.53	288	3.37	223	3.32
5M039	486	3.34	854	3.37	733	3.41	288	3.25	223	3.32
5M043	486	2.73	854	2.75	733	2.77	288	2.66	223	2.80
5M051	486	3.32	854	3.36	733	3.40	288	3.25	223	3.34
5M056	486	3.32	854	3.31	733	3.37	288	3.24	223	3.27
5M062	486	3.14	854	3.20	733	3.24	288	3.07	223	3.16

Table D.46. Item Means by Gender and Ethnicity – Grade 6 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
6M005	533	3.18	969	3.15	818	3.25	354	3.07	252	3.08
6M015	533	2.83	969	2.83	818	2.89	354	2.76	252	2.76
6M021	533	3.02	969	3.16	818	3.18	354	3.05	252	3.04
6M026	533	3.40	969	3.51	818	3.55	354	3.45	252	3.33
6M028	533	3.26	969	3.39	818	3.43	354	3.32	252	3.21
6M035	533	3.15	969	3.28	818	3.30	354	3.23	252	3.15

Table D.46. Item Means by Gender and Ethnicity – Grade 6 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
6M04 3	533	3.26	96 9	3.39	81 8	3.41	35 4	3.25	252	3.29
6M05 3	533	3.43	96 9	3.54	81 8	3.57	35 4	3.48	252	3.41
6M06 1	533	3.20	96 9	3.33	81 8	3.35	35 4	3.23	252	3.21
6M06 4	533	3.33	96 9	3.47	81 8	3.50	35 4	3.33	252	3.38

Table D.47. Item Means by Gender and Ethnicity – Grade 7 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
7M00 3	559	3.14	93 2	3.05	77 9	3.20	33 4	2.91	287	2.93
7M01 2	559	2.63	93 2	2.62	77 9	2.70	33 4	2.51	287	2.55
7M02 3	559	3.09	93 2	3.08	77 9	3.17	33 4	2.88	287	3.06
7M02 4	559	3.25	93 2	3.16	77 9	3.24	33 4	3.09	287	3.21
7M03 0	559	3.04	93 2	3.02	77 9	3.09	33 4	2.89	287	2.96
7M03 9	559	2.50	93 2	2.46	77 9	2.56	33 4	2.32	287	2.39
7M04 3	558	2.95	93 2	2.95	77 8	2.95	33 4	2.81	287	3.04
7M05 4	558	3.53	93 2	3.42	77 8	3.54	33 4	3.31	287	3.44

Table D.47. Item Means by Gender and Ethnicity – Grade 7 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
7M05 7	558	2.81	93 2	2.83	77 8	2.89	33 4	2.66	287	2.80
7M06 6	558	2.95	93 2	2.87	77 8	2.98	33 4	2.76	287	2.85

Table D.48. Item Means by Gender and Ethnicity – Grade 8 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
8M00 1	600	3.36	93 0	3.34	81 3	3.39	40 2	3.28	244	3.31
8M01 0	600	2.99	93 0	3.03	81 3	3.09	40 2	2.85	244	3.03
8M02 1	600	3.47	93 0	3.43	81 3	3.51	40 2	3.35	244	3.39
8M02 5	600	3.54	93 0	3.48	81 3	3.54	40 2	3.40	244	3.54
8M02 9	600	2.96	93 0	3.03	81 3	3.02	40 2	2.95	244	3.00
8M03 7	600	3.02	93 0	3.07	81 3	3.10	40 2	2.93	244	3.07
8M04 5	600	2.84	93 0	2.80	81 3	2.85	40 2	2.73	244	2.86
8M04 9	600	3.48	93 0	3.45	81 3	3.52	40 2	3.36	244	3.46
8M05 3	600	3.28	93 0	3.30	81 3	3.34	40 2	3.17	244	3.34
8M05 8	600	3.16	93 0	3.11	81 3	3.18	40 2	3.03	244	3.09

Table D.49. Item Means by Gender and Ethnicity – Grade 11 Mathematics										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
11M00 2	445	3.19	70 4	3.17	66 8	3.23	28 7	3.14	149	3.07
11M01 7	445	3.01	70 4	2.97	66 8	3.05	28 7	2.90	149	2.83
11M02 4	445	2.72	70 4	2.67	66 8	2.74	28 7	2.60	149	2.63
11M02 7	445	2.79	70 4	2.85	66 8	2.87	28 7	2.77	149	2.74
11M03 3	445	2.96	70 4	2.87	66 8	2.98	28 7	2.78	149	2.79
11M03 7	445	3.14	70 4	3.21	66 8	3.26	28 7	3.07	149	3.15
11M04 2	445	3.16	70 4	3.15	66 8	3.26	28 7	2.96	149	3.08
11M04 9	445	3.42	70 3	3.43	66 7	3.50	28 7	3.31	149	3.34
11M05 9	445	3.32	70 3	3.30	66 7	3.35	28 7	3.26	149	3.28
11M06 1	445	3.36	70 3	3.34	66 7	3.43	28 7	3.23	149	3.25

Table D.50. Item Means by Gender and Ethnicity - Grade 3 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
3R00 1	437	3.14	81 6	3.25	66 7	3.30	25 9	3.10	227	3.05
3R00 8	437	3.37	81 6	3.39	66 7	3.48	25 9	3.25	227	3.24
3R01 1	437	3.40	81 6	3.44	66 7	3.53	25 9	3.24	227	3.29
3R01 8	437	3.17	81 6	3.24	66 7	3.34	25 9	2.98	227	3.05

Table D.50. Item Means by Gender and Ethnicity - Grade 3 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
3R02 2	437	3.23	81 6	3.22	66 7	3.31	25 9	3.08	227	3.12
3R02 9	437	3.23	81 6	3.30	66 7	3.40	25 9	3.03	227	3.12
3R04 4	437	3.41	81 6	3.36	66 7	3.48	25 9	3.18	227	3.29
3R04 6	437	3.16	81 6	3.21	66 7	3.31	25 9	2.99	227	3.00
3R05 1	437	2.86	81 6	2.91	66 7	3.01	25 9	2.75	227	2.72

Table D.51. Item Means by Gender and Ethnicity - Grade 4 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
4R00 4	479	3.39	90 2	3.44	77 6	3.46	29 6	3.38	225	3.32
4R00 8	479	3.41	90 2	3.46	77 6	3.47	29 6	3.44	225	3.37
4R01 1	479	3.44	90 2	3.41	77 6	3.45	29 6	3.31	225	3.44
4R01 8	479	3.32	90 2	3.28	77 6	3.34	29 6	3.18	225	3.25
4R02 2	479	3.23	90 2	3.25	77 6	3.25	29 6	3.14	225	3.28
4R03 5	479	3.14	90 2	3.22	77 6	3.23	29 6	3.14	225	3.16
4R04 0	479	3.15	90 2	3.14	77 6	3.17	29 6	3.10	225	3.12
4R04 9	479	3.38	90 2	3.44	77 6	3.47	29 6	3.32	225	3.35
4R05 4	479	2.68	90 2	2.77	77 6	2.74	29 6	2.64	225	2.78

Table D.52. Item Means by Gender and Ethnicity - Grade 5 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
5R00 2	486	3.36	85 3	3.37	73 2	3.43	28 8	3.31	223	3.32
5R01 2	486	3.34	85 3	3.37	73 2	3.45	28 8	3.24	223	3.30
5R01 3	486	3.23	85 3	3.25	73 2	3.30	28 8	3.13	223	3.24
5R02 2	486	3.16	85 3	3.11	73 2	3.19	28 8	3.04	223	3.13
5R02 5	486	3.47	85 3	3.48	73 2	3.57	28 8	3.29	223	3.44
5R03 2	486	3.13	85 3	3.06	73 2	3.18	28 8	2.92	223	3.05
5R04 3	486	3.33	85 3	3.27	73 2	3.35	28 8	3.16	223	3.31
5R05 6	486	3.33	85 3	3.29	73 2	3.40	28 8	3.16	223	3.26
5R06 2	486	3.00	85 3	2.94	73 2	3.02	28 8	2.88	223	2.88

Table D.53. Item Means by Gender and Ethnicity - Grade 6 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
6R00 3	533	3.38	96 8	3.44	81 8	3.50	35 4	3.38	251	3.29
6R00 8	533	3.25	96 8	3.32	81 8	3.36	35 4	3.20	251	3.30
6R01 5	533	3.06	96 8	3.18	81 8	3.25	35 4	3.07	251	3.00
6R02 1	533	3.11	96 8	3.16	81 8	3.25	35 4	3.01	251	3.03
6R03 7	533	2.92	96 8	2.91	81 8	2.99	35 4	2.83	251	2.82

Table D.53. Item Means by Gender and Ethnicity - Grade 6 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
6R040	533	3.06	968	3.19	818	3.25	354	3.01	251	3.03
6R047	533	3.28	968	3.42	818	3.50	354	3.25	251	3.21
6R050	533	3.35	967	3.45	818	3.49	353	3.39	251	3.30
6R063	533	3.09	967	3.11	818	3.22	353	3.03	251	2.94

Table D.54. Item Means by Gender and Ethnicity - Grade 7 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
7R00 2	560	3.44	93 3	3.34	78 0	3.47	33 4	3.27	287	3.23
7R00 5	560	3.39	93 3	3.29	78 0	3.40	33 4	3.19	287	3.30
7R01 1	560	3.33	93 3	3.27	78 0	3.37	33 4	3.15	287	3.20
7R02 1	560	3.47	93 3	3.35	78 0	3.48	33 4	3.28	287	3.32
7R02 9	560	3.31	93 3	3.21	78 0	3.36	33 4	3.06	287	3.17
7R04 3	560	3.15	93 3	3.00	78 0	3.12	33 4	3.02	287	2.98
7R04 6	559	3.36	93 3	3.27	77 9	3.41	33 4	3.11	287	3.25
7R05 4	559	3.23	93 3	3.11	77 9	3.23	33 4	2.96	287	3.16
7R05 7	559	2.96	93 3	2.85	77 9	2.93	33 4	2.72	287	2.94

Table D.55. Item Means by Gender and Ethnicity - Grade 8 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
8R00 1	599	3.58	93 2	3.56	81 2	3.64	40 2	3.48	246	3.51
8R00 8	599	3.41	93 2	3.33	81 2	3.42	40 2	3.23	246	3.36
8R01 9	598	3.03	93 2	3.03	81 1	3.10	40 2	2.84	246	3.07
8R02 2	599	3.15	93 2	3.09	81 2	3.14	40 2	3.01	246	3.17
8R03 2	599	3.51	93 2	3.49	81 2	3.57	40 2	3.43	246	3.41

Table D.55. Item Means by Gender and Ethnicity - Grade 8 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
8R03 9	599	3.35	93 2	3.28	81 2	3.36	40 2	3.19	246	3.29
8R04 4	598	3.29	93 2	3.28	81 1	3.36	40 2	3.13	246	3.28
8R04 6	598	3.30	93 2	3.23	81 1	3.32	40 2	3.17	246	3.24
8R05 1	598	3.33	93 2	3.24	81 1	3.33	40 2	3.21	246	3.24

Table D.56. Item Means by Gender and Ethnicity - Grade 11 Reading										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
11R003	446	3.61	705	3.63	670	3.69	287	3.55	149	3.50
11R007	446	3.41	705	3.42	670	3.50	287	3.29	149	3.32
11R014	446	3.56	705	3.56	670	3.63	287	3.43	149	3.49
11R020	446	3.58	705	3.58	670	3.65	287	3.49	149	3.50
11R024	446	3.52	705	3.53	670	3.61	287	3.45	149	3.40
11R035	446	3.46	705	3.42	670	3.51	287	3.33	149	3.35
11R038	446	3.39	705	3.33	670	3.45	287	3.22	149	3.23
11R044	446	3.39	705	3.37	670	3.47	287	3.20	149	3.30
11R051	446	3.50	705	3.43	670	3.54	287	3.31	149	3.41

Table D.57. Item Means by Gender and Ethnicity - Grade 4 Science										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
4S008	477	3.08	900	3.13	775	3.13	294	3.08	224	3.12
4S014	477	3.35	900	3.38	775	3.41	294	3.33	224	3.29
4S026	477	3.11	900	3.15	775	3.16	294	3.06	224	3.12
4S035	477	3.33	900	3.41	775	3.41	294	3.33	224	3.35

Table D.57. Item Means by Gender and Ethnicity - Grade 4 Science										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
4S042	477	3.29	900	3.41	775	3.39	294	3.35	224	3.32
4S057	477	3.34	900	3.32	775	3.37	294	3.24	224	3.27

Table D.58. Item Means by Gender and Ethnicity - Grade 7 Science										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
7S003	557	3.45	931	3.38	777	3.49	333	3.32	287	3.31
7S015	557	3.55	931	3.49	777	3.58	333	3.45	287	3.41
7S024	557	3.30	931	3.27	777	3.41	333	3.06	287	3.21
7S037	557	3.06	931	3.07	777	3.13	333	2.97	287	2.99
7S045	557	3.40	931	3.28	777	3.39	333	3.21	287	3.27
7S052	557	3.26	931	3.20	777	3.28	333	3.17	287	3.10

Table D.59. Item Means by Gender and Ethnicity - Grade 11 Science										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
11S008	444	3.27	703	3.31	667	3.36	286	3.24	149	3.17
11S016	444	3.05	703	3.18	667	3.23	286	3.03	149	2.94
11S023	444	3.44	703	3.48	667	3.52	286	3.38	149	3.39

Table D.59. Item Means by Gender and Ethnicity - Grade 11 Science										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
11S03 5	444	3.02	70 3	3.12	66 7	3.18	28 6	2.91	149	2.99
11S04 2	444	3.17	70 3	3.30	66 7	3.34	28 6	3.14	149	3.11
11S04 5	444	3.64	70 3	3.61	66 7	3.69	28 6	3.56	149	3.47

Table D.60. Item Means by Gender and Ethnicity - Grade 5 Writing										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
5W02 1	486	3.41	85 2	3.38	73 3	3.47	28 7	3.24	222	3.31
5W02 9	486	3.13	85 2	3.19	73 3	3.23	28 7	3.02	222	3.14
5W10 0	486	3.14	85 2	3.09	73 3	3.17	28 7	3.00	222	3.10
5W12 3	486	3.08	85 2	3.10	73 3	3.17	28 7	2.95	222	3.02
5W13 0	486	3.05	85 2	3.05	73 3	3.14	28 7	2.88	222	2.95

Table D.61. Item Means by Gender and Ethnicity - Grade 6 Writing										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
6W00 3	532	2.99	96 9	2.99	81 7	3.07	35 4	2.90	252	2.92
6W01 2	532	2.86	96 9	2.87	81 7	2.91	35 4	2.84	252	2.79
6W01 8	532	2.95	96 9	2.97	81 7	3.07	35 4	2.91	252	2.77

Table D.61. Item Means by Gender and Ethnicity - Grade 6 Writing										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
6W021	532	3.38	969	3.51	817	3.56	354	3.38	252	3.32
6W030	532	2.64	969	2.64	817	2.68	354	2.65	252	2.56

Table D.62. Item Means by Gender and Ethnicity - Grade 8 Writing										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
8W038	598	3.28	928	3.22	811	3.34	399	3.10	245	3.15
8W043	599	3.26	930	3.26	811	3.35	402	3.11	245	3.24
8W052	599	3.47	930	3.41	811	3.52	402	3.31	245	3.41
8W054	599	3.29	930	3.15	811	3.27	402	3.05	245	3.24
8W060	599	3.41	930	3.33	811	3.46	402	3.17	245	3.34

Table D.63. Item Means by Gender and Ethnicity - Grade 11 Writing										
	Female		Male		White		Black		Hispanic	
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
9W028	446	3.34	704	3.36	669	3.43	287	3.24	149	3.23
9W101	446	3.35	704	3.29	669	3.42	287	3.19	149	3.13
9W107	446	3.42	704	3.41	669	3.47	287	3.40	149	3.23
9W110	446	3.32	704	3.31	669	3.36	287	3.26	149	3.18
9W117	446	3.39	704	3.42	669	3.49	287	3.32	149	3.27

Table D.64. IRT Item Analysis Results – Grade 3 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
3M002	1250	1.03	-1.30	-0.03	-1.76	-2.89	-0.80	0.24	1.60

Table D.64. IRT Item Analysis Results – Grade 3 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
3M01 3	125 0	0.82	- 0.59	0.39	0.33	- 1.98	0.24	0.9 0	2.08
3M02 2	125 0	1.21	- 0.45	1.15	1.38	- 1.29	0.65	1.5 2	2.29
3M02 9	125 0	1.10	- 0.58	0.93	1.76	- 1.47	0.66	1.4 1	2.43
3M03 6	125 0	0.98	- 0.63	0.49	- 0.44	- 2.00	0.00	0.7 3	1.89
3M03 8	125 0	1.06	- 1.09	0.05	0.09	- 2.35	0.05	0.7 9	1.90
3M04 7	125 0	0.96	- 0.56	0.80	0.54	- 1.72	0.38	1.1 9	2.12
3M04 9	125 0	1.10	- 1.31	- 0.06	- 1.01	- 2.69	- 0.42	0.5 4	1.68
3M05 6	125 0	1.11	- 0.46	0.55	1.08	- 1.54	0.41	1.2 6	2.19
3M06 1	124 9	0.95	- 0.67	0.64	0.77	- 1.79	0.32	1.1 6	2.18

Table D.65. IRT Item Analysis Results – Grade 4 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
4M00 2	138 1	0.91	- 1.30	0.64	- 0.33	- 2.20	- 0.03	0.8 0	2.14
4M00 5	138 1	0.90	- 0.91	0.45	- 0.46	- 2.11	0.01	0.7 8	2.10
4M02 4	138 1	1.23	- 0.46	1.35	1.52	- 0.93	0.79	1.4 7	2.65
4M03 1	138 1	0.93	- 0.67	0.78	0.14	- 1.69	0.28	1.0 8	2.26

Table D.65. IRT Item Analysis Results – Grade 4 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
4M03 6	138 1	0.95	- 1.15	0.52	- 0.34	- 2.12	0.00	0.7 6	2.12
4M04 0	138 1	1.03	- 1.31	0.39	0.36	- 2.26	0.24	1.0 6	2.21
4M04 5	138 1	1.25	- 0.55	1.13	1.80	- 1.16	0.79	1.7 3	2.57
4M04 9	138 1	1.16	- 1.12	0.03	0.41	- 2.06	0.09	1.2 0	2.12
4M05 6	138 1	0.91	- 0.68	- 0.04	- 0.56	- 2.16	- 0.16	0.8 0	2.02
4M06 1	138 1	0.80	- 0.53	0.58	0.34	- 1.70	0.28	1.0 4	2.32

Table D.66. IRT Item Analysis Results – Grade 5 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
5M00 2	134 0	1.06	- 0.92	0.6 4	0.87	- 1.60	0.40	1.1 9	1.96
5M01 2	134 0	1.10	- 0.37	1.1 2	1.78	- 0.95	0.66	1.5 9	2.22
5M01 6	134 0	0.98	- 0.30	0.8 0	1.02	- 1.10	0.49	1.2 5	2.10
5M02 6	134 0	1.10	- 1.42	0.4 6	0.25	- 2.06	0.15	0.9 8	1.79
5M03 5	134 0	1.01	- 1.37	0.2 8	- 0.80	- 2.31	- 0.27	0.6 7	1.62
5M03 9	134 0	0.89	- 0.99	0.2 7	- 0.45	- 2.05	- 0.16	0.6 7	1.72
5M04 3	134 0	1.14	- 0.59	1.1 1	1.25	- 1.09	0.61	1.3 5	2.12

Table D.66. IRT Item Analysis Results – Grade 5 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
5M05 1	134 0	0.97	- 1.37	0.3 0	- 0.28	- 2.24	- 0.09	0.7 5	1.72
5M05 6	134 0	0.95	- 0.87	0.2 7	- 0.29	- 1.90	- 0.04	0.8 6	1.71
5M06 2	134 0	1.01	- 1.04	0.3 2	0.35	- 1.95	0.22	0.9 5	1.83

Table D.67. IRT Item Analysis Results – Grade 6 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
6M00 5	150 2	1.25	- 1.02	0.6 4	1.03	- 1.74	0.39	1.4 6	2.58
6M01 5	150 2	1.22	- 0.70	1.2 0	2.18	- 1.19	0.76	1.7 4	3.03
6M02 1	150 2	1.09	- 0.83	0.6 3	1.23	- 1.74	0.51	1.4 3	2.69
6M02 6	150 2	1.06	- 1.27	0.2 6	- 0.31	- 2.40	- 0.04	0.7 1	2.32
6M02 8	150 2	0.94	- 0.96	0.5 2	0.11	- 2.13	0.15	1.0 4	2.44
6M03 5	150 2	0.84	- 0.83	0.5 3	0.63	- 1.93	0.19	1.1 5	2.61
6M04 3	150 2	1.12	- 1.16	0.3 1	0.34	- 2.21	0.13	1.2 7	2.39
6M05 3	150 2	0.96	- 1.52	0.0 3	- 0.29	- 2.71	- 0.23	0.7 4	2.29
6M06 1	150 2	0.78	- 1.02	0.5 7	0.45	- 2.19	0.23	0.9 9	2.58
6M06 4	150 2	0.86	- 0.91	0.5 8	- 0.40	- 2.27	0.06	0.9 4	2.35

Table D.68. IRT Item Analysis Results – Grade 7 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
7M003	1491	1.07	- 1.12	0.2 1	0.34	- 1.82	- 0.11	0.9 4	1.56
7M012	1491	1.17	- 1.02	0.9 5	1.54	- 1.41	0.46	1.2 3	1.87
7M023	1491	0.89	- 0.86	0.2 3	0.21	- 1.87	0.02	0.7 3	1.64
7M024	1491	1.05	- 1.40	0.3 8	- 0.16	- 2.00	- 0.15	0.6 5	1.52
7M030	1491	0.93	- 0.95	0.4 3	0.32	- 1.88	0.15	0.8 1	1.66
7M039	1491	1.22	- 0.79	1.3 0	1.75	- 1.15	0.61	1.3 2	1.97
7M043	1490	0.88	- 0.90	0.4 7	0.57	- 1.82	0.18	0.9 1	1.74
7M054	1490	0.96	- 1.56	0.0 7	- 1.13	- 2.66	- 0.44	0.3 6	1.33
7M057	1490	1.09	- 0.91	0.8 0	0.74	- 1.56	0.31	1.1 4	1.70
7M066	1490	0.92	- 0.58	0.5 6	0.50	- 1.53	0.21	0.9 5	1.74

Table D.69. IRT Item Analysis Results – Grade 8 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
8M001	1530	1.16	- 1.28	0.4 7	- 0.04	- 1.93	- 0.02	1.0 3	2.12
8M010	1530	0.99	- 0.70	1.0 7	0.85	- 1.36	0.43	1.4 3	2.45
8M021	1530	1.00	- 1.45	0.1 7	- 0.27	- 2.25	- 0.24	0.7 1	2.09

Table D.69. IRT Item Analysis Results – Grade 8 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
8M02 5	153 0	1.02	- 1.60	0.1 4	- 0.57	- 2.46	- 0.33	0.6 8	2.01
8M02 9	153 0	1.16	- 0.85	0.7 3	1.28	- 1.49	0.48	1.5 1	2.42
8M03 7	153 0	0.97	- 0.79	0.8 5	0.92	- 1.57	0.43	1.3 8	2.43
8M04 5	153 0	1.27	- 0.64	0.8 9	2.02	- 1.23	0.70	1.8 1	2.51
8M04 9	153 0	0.91	- 1.15	0.1 7	- 0.48	- 2.21	- 0.27	0.7 3	2.06
8M05 3	153 0	0.84	- 1.02	0.6 2	0.03	- 1.96	0.09	0.9 0	2.24
8M05 8	153 0	0.94	- 0.71	0.6 9	0.66	- 1.52	0.26	1.2 4	2.38

Table D.70. IRT Item Analysis Results – Grade 11 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
11M00 2	114 9	1.00	- 0.89	0.27	0.31	- 2.09	0.15	0.9 9	1.77
11M01 7	114 9	1.06	- 0.80	0.28	1.10	- 1.89	0.37	1.1 8	1.90
11M02 4	114 9	1.18	- 0.70	0.89	1.80	- 1.42	0.60	1.6 0	1.95
11M02 7	114 9	1.04	- 0.78	0.83	1.26	- 1.53	0.44	1.3 6	2.01
11M03 3	114 9	0.92	- 0.55	0.63	1.01	- 1.62	0.45	1.2 0	2.01
11M03 7	114 9	0.96	- 0.85	0.23	0.31	- 2.18	0.22	0.9 6	1.78

Table D.70. IRT Item Analysis Results – Grade 11 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
11M04 2	114 9	0.94	- 1.01	0.31	0.41	- 2.27	0.28	0.9 2	1.82
11M04 9	114 8	1.18	- 1.49	- 0.20	- 0.30	- 2.71	- 0.24	0.8 7	1.55
11M05 9	114 8	0.98	- 0.71	- 0.04	- 0.08	- 2.29	0.08	0.9 8	1.64
11M06 1	114 8	0.95	- 1.06	- 0.03	- 0.16	- 2.53	- 0.07	0.8 9	1.64

Table D.71. IRT Item Analysis Results - Grade 3 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
3R00 1	125 3	1.09	- 1.13	0.35	0.89	- 2.25	0.26	1.3 1	2.47
3R00 8	125 3	0.90	- 1.13	- 0.18	0.29	- 2.63	- 0.15	0.9 9	2.34
3R01 1	125 3	0.95	- 1.10	0.08	- 0.16	- 2.45	- 0.31	0.8 7	2.28
3R01 8	125 3	1.00	- 0.84	0.56	0.60	- 2.16	0.34	1.3 5	2.45
3R02 2	125 3	0.95	- 1.17	0.43	0.77	- 2.29	0.13	1.2 3	2.50
3R02 9	125 3	1.03	- 1.25	0.22	0.70	- 2.34	- 0.06	1.2 8	2.43
3R04 4	125 3	1.07	- 1.12	0.26	0.01	- 2.41	0.06	0.8 0	2.33
3R04 6	125 3	1.09	- 0.61	0.24	0.91	- 1.92	0.21	1.4 2	2.47
3R05 1	125 3	1.07	- 0.24	1.33	1.29	- 1.39	0.85	1.7 2	2.71

Table D.72. IRT Item Analysis Results - Grade 4 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
4R004	1381	0.99	-1.27	0.49	-0.23	-2.24	-0.09	0.95	2.11
4R008	1381	0.88	-1.55	0.01	0.04	-2.65	-0.25	0.91	2.13
4R011	1381	0.90	-1.21	0.45	-0.21	-2.30	-0.11	0.90	2.13
4R018	1381	1.08	-0.76	0.30	0.42	-1.93	0.24	1.31	2.16
4R022	1381	1.02	-0.71	0.51	0.49	-1.83	0.32	1.26	2.24
4R035	1381	1.12	-0.89	0.55	0.77	-1.86	0.43	1.33	2.27
4R040	1381	1.02	-0.97	0.79	0.84	-1.92	0.45	1.50	2.30
4R049	1381	0.95	-0.96	0.19	-0.08	-2.26	-0.02	0.96	2.12
4R054	1381	1.30	-0.26	1.30	1.95	-1.07	1.08	1.91	2.47

Table D.73. IRT Item Analysis Results - Grade 5 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
5R00 2	133 9	1.04	- 1.03	0.2 3	0.16	- 2.01	- 0.07	1.0 3	2.17
5R01 2	133 9	1.02	- 1.22	0.1 2	0.34	- 2.28	0.05	1.0 9	2.17
5R01 3	133 9	0.94	- 1.04	0.4 9	0.60	- 1.99	0.23	1.2 7	2.27
5R02 2	133 9	1.02	- 0.82	0.6 3	0.95	- 1.71	0.43	1.4 0	2.35
5R02 5	133 9	0.91	- 1.42	0.0 3	- 0.22	- 2.49	- 0.28	0.7 6	2.10
5R03 2	133 9	1.17	- 0.99	0.6 4	1.22	- 1.74	0.52	1.5 4	2.33
5R04 3	133 9	0.92	- 0.96	0.2 6	0.50	- 2.09	0.12	1.2 4	2.23
5R05 6	133 9	1.08	- 0.92	0.1 0	0.55	- 1.95	0.08	1.2 4	2.19
5R06 2	133 9	1.11	- 0.53	1.0 5	1.27	- 1.28	0.68	1.6 1	2.44

Table D.74. IRT Item Analysis Results - Grade 6 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
6R00 3	150 1	1.12	- 1.56	- 0.02	0.09	- 2.29	- 0.20	0.9 0	2.01
6R00 8	150 1	1.01	- 1.21	0.22	0.46	- 2.20	0.21	1.0 3	2.12
6R01 5	150 1	0.79	- 0.40	0.71	0.48	- 1.53	0.38	1.1 4	2.27
6R02 1	150 1	1.03	- 0.96	0.34	1.00	- 1.80	0.29	1.2 9	2.23
6R03 7	150 1	1.30	- 0.75	0.85	1.54	- 1.32	0.68	1.5 9	2.29

Table D.74. IRT Item Analysis Results - Grade 6 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
6R04 0	150 1	1.09	- 0.75	0.54	0.77	- 1.56	0.34	1.2 9	2.20
6R04 7	150 1	0.98	- 1.17	0.08	0.16	- 2.14	- 0.05	0.8 6	2.08
6R05 0	150 0	0.84	- 1.20	0.10	- 0.10	- 2.44	- 0.07	0.8 7	2.02
6R06 3	150 0	1.02	- 0.69	0.88	0.60	- 1.59	0.50	1.3 6	2.21

Table D.75. IRT Item Analysis Results - Grade 7 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
7R00 2	149 3	1.09	- 1.51	0.0 2	0.3 0	- 2.14	- 0.23	1.0 2	2.27
7R00 5	149 3	1.00	- 1.20	0.3 2	0.2 5	- 1.95	- 0.07	1.0 2	2.33
7R01 1	149 3	0.94	- 1.10	0.3 4	0.4 1	- 2.03	0.12	1.0 4	2.38
7R02 1	149 3	1.16	- 1.68	0.2 8	0.0 6	- 2.15	- 0.17	0.9 3	2.25
7R02 9	149 3	0.93	- 0.76	0.4 5	0.4 1	- 1.74	0.13	1.0 8	2.41
7R04 3	149 3	1.04	- 0.78	1.0 0	0.8 8	- 1.54	0.56	1.2 8	2.58
7R04 6	149 2	0.92	- 1.17	0.3 8	0.3 3	- 2.01	- 0.02	1.0 6	2.37
7R05 4	149 2	0.94	- 0.80	0.5 2	0.8 6	- 1.74	0.31	1.3 2	2.50
7R05 7	149 2	1.15	- 0.39	0.9 3	1.6 6	- 1.21	0.74	1.6 6	2.70

Table D.76. IRT Item Analysis Results - Grade 8 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
8R00 1	153 1	1.05	- 2.07	0.1 7	- 0.46	- 2.65	- 0.47	0.7 9	2.33
8R00 8	153 1	1.13	- 1.12	0.2 8	0.54	- 1.94	0.20	1.1 6	2.54
8R01 9	153 0	1.01	- 0.74	0.9 5	1.65	- 1.46	0.63	1.7 5	2.89
8R02 2	153 1	1.15	- 0.90	0.7 1	1.46	- 1.52	0.52	1.6 3	2.77
8R03 2	153 1	1.00	- 1.40	0.2 4	- 0.18	- 2.30	- 0.19	0.9 7	2.40

Table D.76. IRT Item Analysis Results - Grade 8 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
8R03 9	153 1	1.10	- 1.47	0.4 1	0.86	- 2.16	0.22	1.4 0	2.59
8R04 4	153 0	0.96	- 0.96	0.5 4	0.71	- 1.78	0.16	1.3 5	2.64
8R04 6	153 0	0.86	- 0.85	0.5 3	0.80	- 1.89	0.28	1.4 5	2.66
8R05 1	153 0	0.94	- 0.93	0.4 6	0.80	- 1.75	0.14	1.3 8	2.65

Table D.77. IRT Item Analysis Results - Grade 11 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
11R003	1151	0.93	- 1.63	0.1 8	- 0.18	- 3.03	- 0.15	1.0 7	2.80
11R007	1151	1.06	- 0.74	0.4 1	0.92	- 2.14	0.41	1.6 7	3.01
11R014	1151	0.96	- 1.01	0.1 5	0.22	- 2.56	0.06	1.1 8	2.89
11R020	1151	0.95	- 1.22	0.0 7	0.13	- 2.79	0.03	1.1 0	2.87
11R024	1151	0.82	- 0.70	0.2 3	0.27	- 2.38	0.07	1.2 6	2.92
11R035	1151	0.97	- 0.80	0.7 4	0.60	- 2.19	0.52	1.4 1	3.02
11R038	1151	1.23	- 0.64	0.4 2	1.29	- 1.88	0.58	1.7 3	3.07
11R044	1151	1.08	- 0.40	0.5 6	0.96	- 1.82	0.56	1.6 7	3.05
11R051	1151	1.13	- 0.95	0.3 4	0.77	- 2.23	0.31	1.5 2	2.98

Table D.78. IRT Item Analysis Results - Grade 4 Science									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
4S008	1377	1.13	- 0.75	0.8 7	1.11	- 1.75	0.62	1.5 6	2.61
4S014	1377	0.89	- 1.29	0.4 8	0.13	- 2.40	- 0.04	1.1 8	2.41
4S026	1377	0.92	- 0.73	1.0 0	0.85	- 1.71	0.45	1.4 2	2.65
4S035	1377	1.06	- 1.47	0.3 8	0.21	- 2.32	- 0.07	0.9 8	2.42

Table D.78. IRT Item Analysis Results - Grade 4 Science									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
4S04 2	137 7	1.03	- 1.20	0.7 6	- 0.12	- 2.19	0.07	1.0 6	2.39
4S05 7	137 7	1.14	- 1.01	0.4 7	0.30	- 1.98	0.10	1.1 4	2.43

Table D.79. IRT Item Analysis Results - Grade 7 Science									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
7S00 3	148 8	0.97	- 1.28	0.3 3	0.26	- 2.12	- 0.07	1.0 1	2.49
7S01 5	148 8	1.04	- 2.35	0.2 1	- 0.06	- 2.72	- 0.42	0.8 7	2.38
7S02 4	148 8	0.98	- 0.75	0.6 6	0.53	- 1.73	0.21	1.4 4	2.56
7S03 7	148 8	1.24	- 0.67	0.8 4	1.55	- 1.43	0.73	1.7 0	2.72
7S04 5	148 8	0.99	- 1.07	0.6 2	0.44	- 1.91	0.12	1.3 3	2.53
7S05 2	148 8	0.97	- 0.82	0.6 9	0.89	- 1.78	0.45	1.3 7	2.66

Table D.80. IRT Item Analysis Results - Grade 11 Science									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
11S00 8	114 7	1.07	- 0.39	0.04	0.69	- 1.72	0.34	1.2 0	2.37
11S01 6	114 7	1.21	- 0.69	0.37	1.38	- 1.58	0.50	1.5 0	2.49
11S02 3	114 7	0.93	- 1.04	- 0.03	0.06	- 2.40	0.00	0.9 8	2.23

Table D.80. IRT Item Analysis Results - Grade 11 Science									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
11S03 5	114 7	1.06	- 0.31	0.82	1.04	- 1.36	0.69	1.5 2	2.52
11S04 2	114 7	1.09	- 0.64	0.29	0.84	- 1.86	0.48	1.3 7	2.38
11S04 5	114 7	0.93	- 1.45	- 0.32	- 0.66	- 2.94	- 0.35	0.5 7	2.09

Table D.81. IRT Item Analysis Results - Grade 5 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
5W02 1	133 8	1.01	- 1.30	0.2 6	0.0 5	- 2.36	- 0.15	0.8 8	2.32
5W02 9	133 8	1.05	- 1.68	0.8 0	0.9 8	- 2.35	0.23	1.4 1	2.54
5W10 0	133 8	1.05	- 0.85	1.0 4	0.7 8	- 1.77	0.43	1.3 8	2.57
5W12 3	133 8	1.03	- 1.00	0.6 1	1.3 1	- 1.93	0.29	1.5 5	2.61
5W13 0	133 8	1.03	- 0.55	0.8 7	1.0 8	- 1.65	0.50	1.5 5	2.61

Table D.82. IRT Item Analysis Results - Grade 6 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
6W00 3	150 1	0.93	- 1.04	0.37	0.86	- 2.20	0.09	1.1 2	2.24
6W01 2	150 1	1.09	- 1.04	0.61	1.27	- 1.90	0.19	1.2 0	2.37
6W01 8	150 1	0.88	- 0.80	0.36	0.87	- 2.06	0.04	1.2 1	2.25

Table D.82. IRT Item Analysis Results - Grade 6 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
6W02 1	150 1	1.15	- 2.48	- 0.49	- 0.61	- 3.28	- 0.83	0.4 6	1.73
6W03 0	150 1	1.08	- 0.94	1.08	1.97	- 1.73	0.52	1.5 0	2.64

Table D.83. IRT Item Analysis Results - Grade 8 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
8W038	1526	1.16	- 1.32	0.1 1	1.17	- 2.46	0.18	1.2 2	2.76
8W043	1529	1.01	- 1.30	0.3 4	0.86	- 2.40	0.04	1.1 8	2.74
8W052	1529	0.98	- 1.73	0.3 2	- 0.17	- 2.73	- 0.35	0.7 2	2.52
8W054	1529	1.05	- 1.32	0.5 6	1.01	- 2.26	0.11	1.2 8	2.80
8W060	1529	0.95	- 1.56	0.2 8	0.31	- 2.53	- 0.32	1.0 3	2.61

Table D.84. IRT Item Analysis Results - Grade 11 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
9W028	1150	1.12	- 1.66	0.18	0.9 0	- 2.59	0.00	1.1 8	2.69
9W101	1150	0.98	- 1.14	0.32	0.8 5	- 2.39	0.14	1.2 7	2.71
9W107	1150	0.90	- 1.27	- 0.03	0.5 3	- 2.69	- 0.12	1.0 5	2.63
9W110	1150	1.18	- 1.30	0.08	1.1 2	- 2.34	- 0.06	1.4 2	2.69
9W117	1150	0.99	- 1.12	- 0.01	0.5 1	- 2.53	0.02	0.9 8	2.62

Figure D.1. Test Information Function – Grade 3 Mathematics

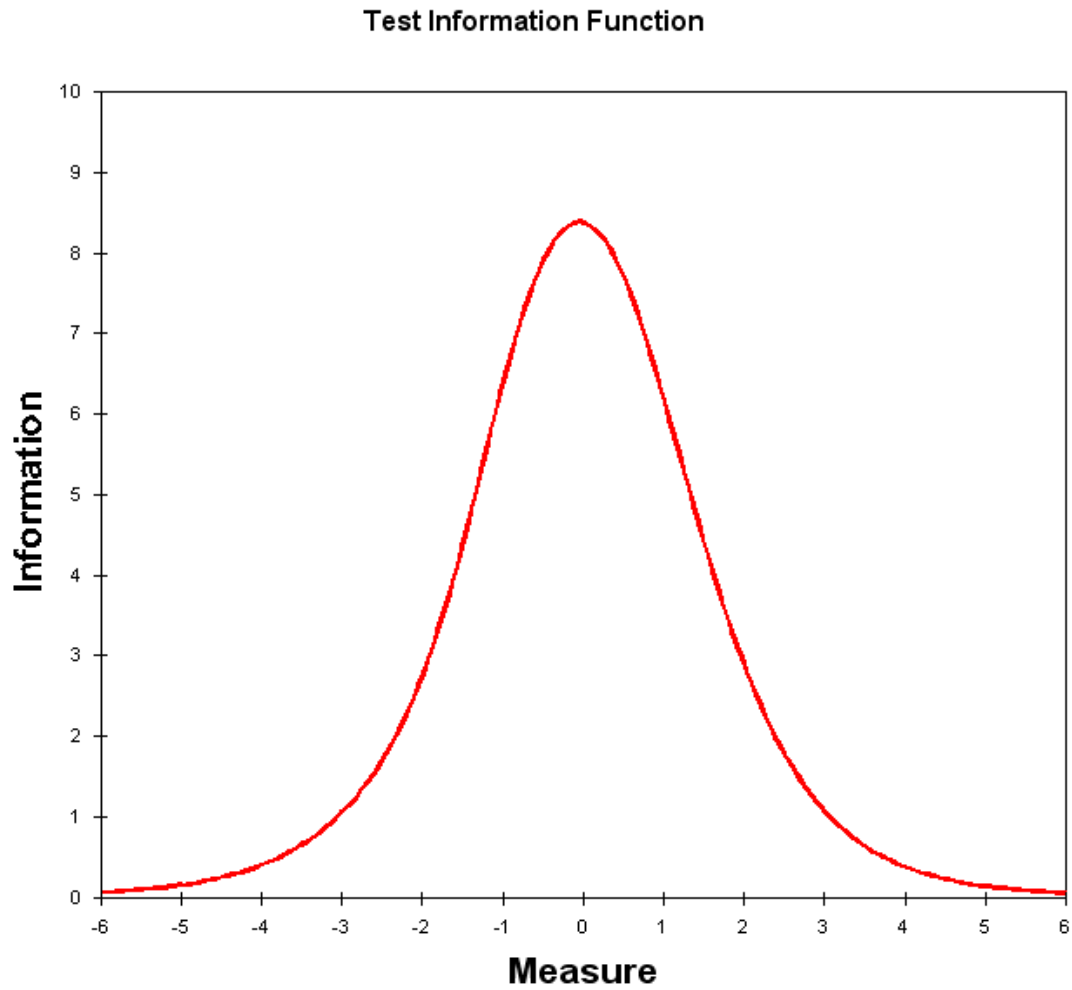


Figure D.2. Test Information Function – Grade 4 Mathematics

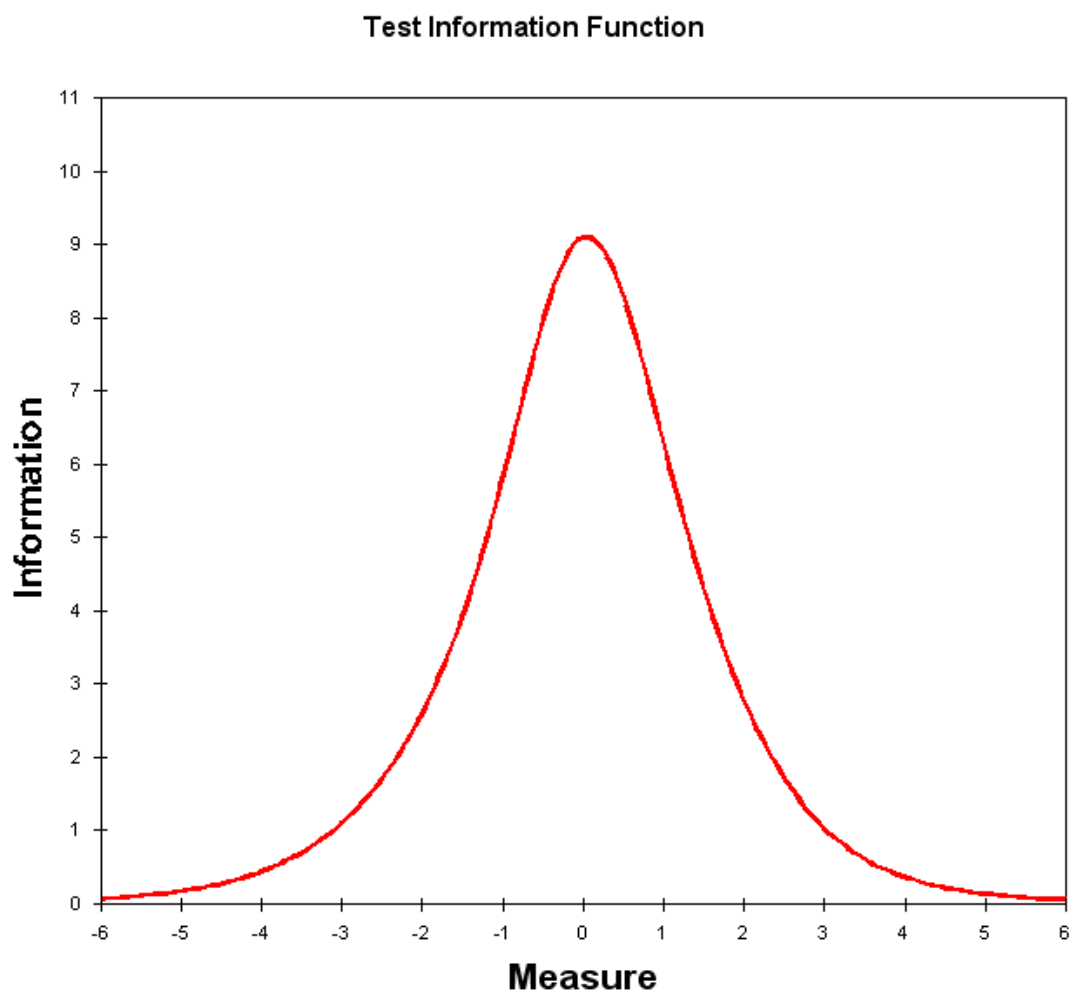


Figure D.3. Test Information Function – Grade 5 Mathematics

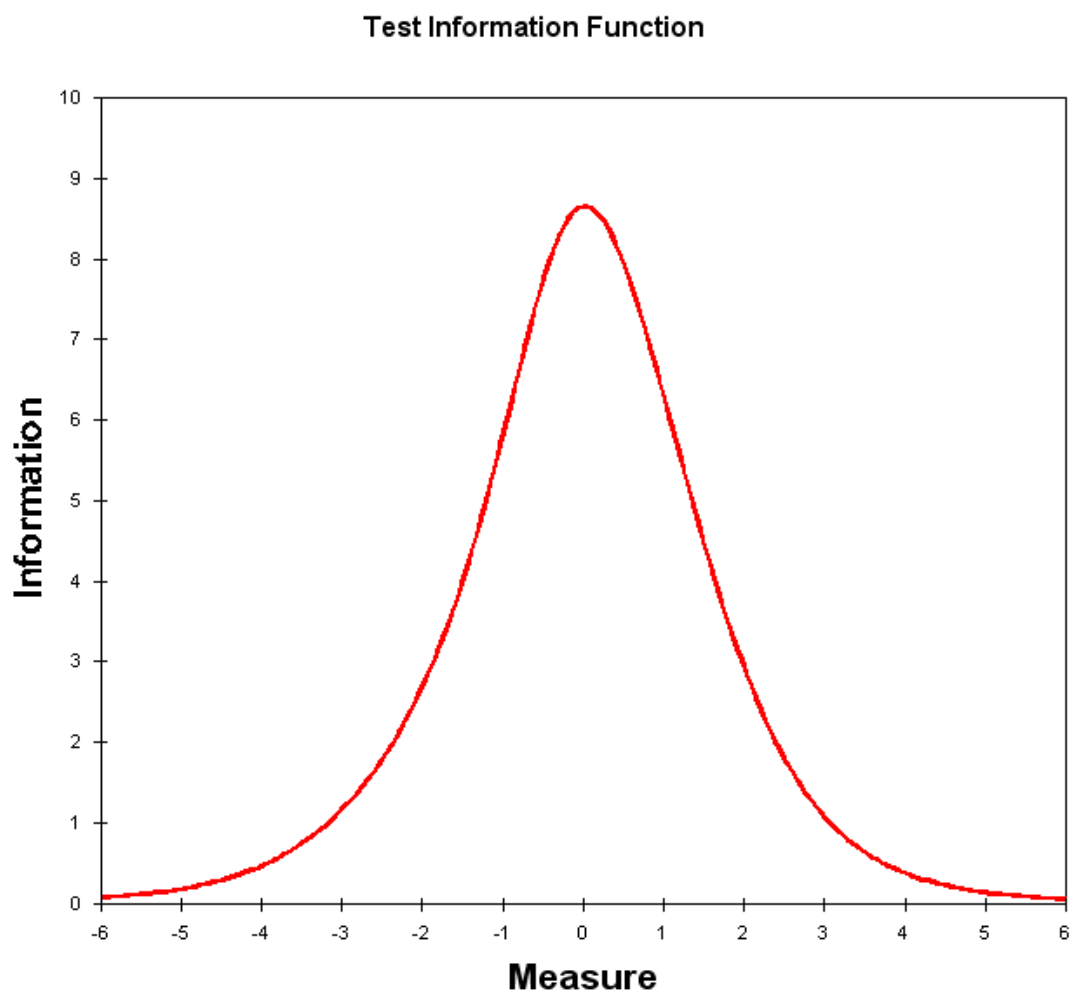


Figure D.4. Test Information Function – Grade 6 Mathematics

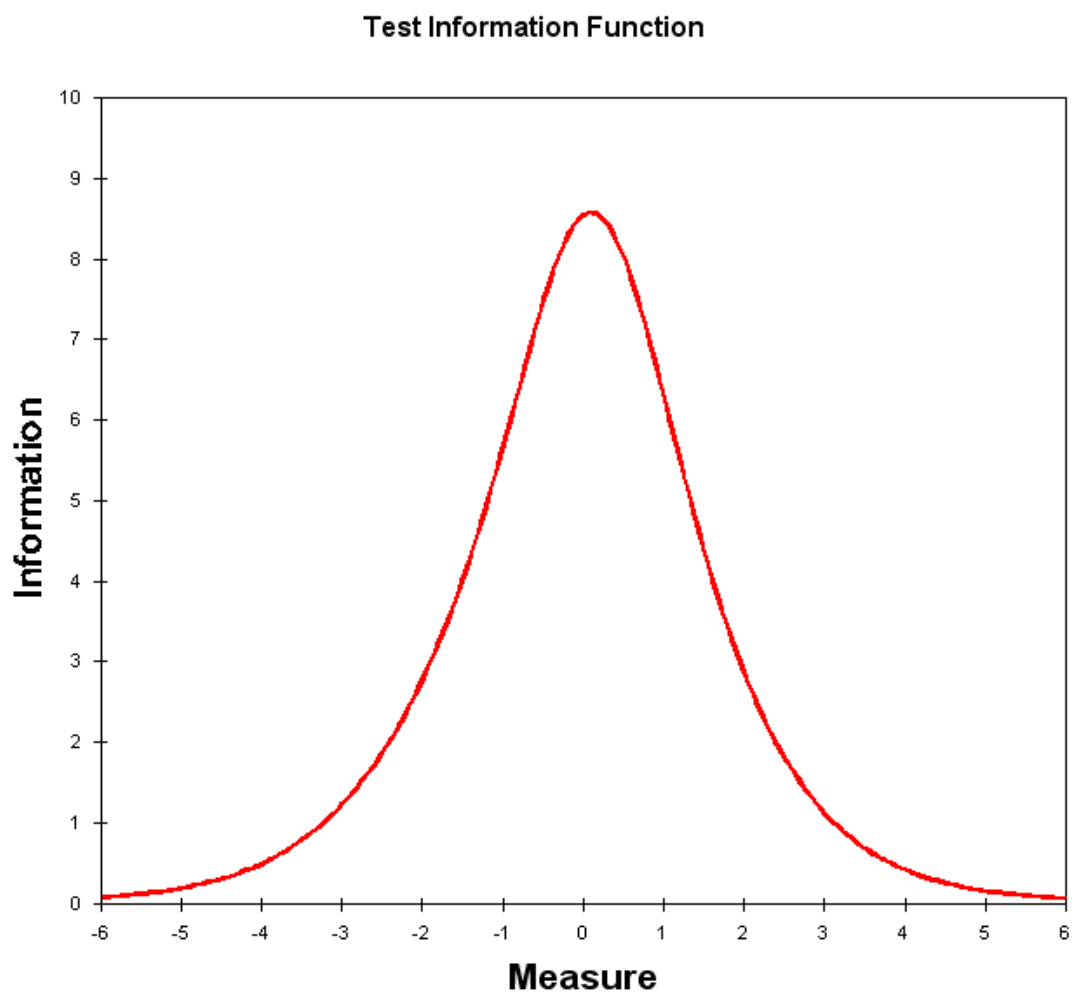


Figure D.5. Test Information Function – Grade 7 Mathematics

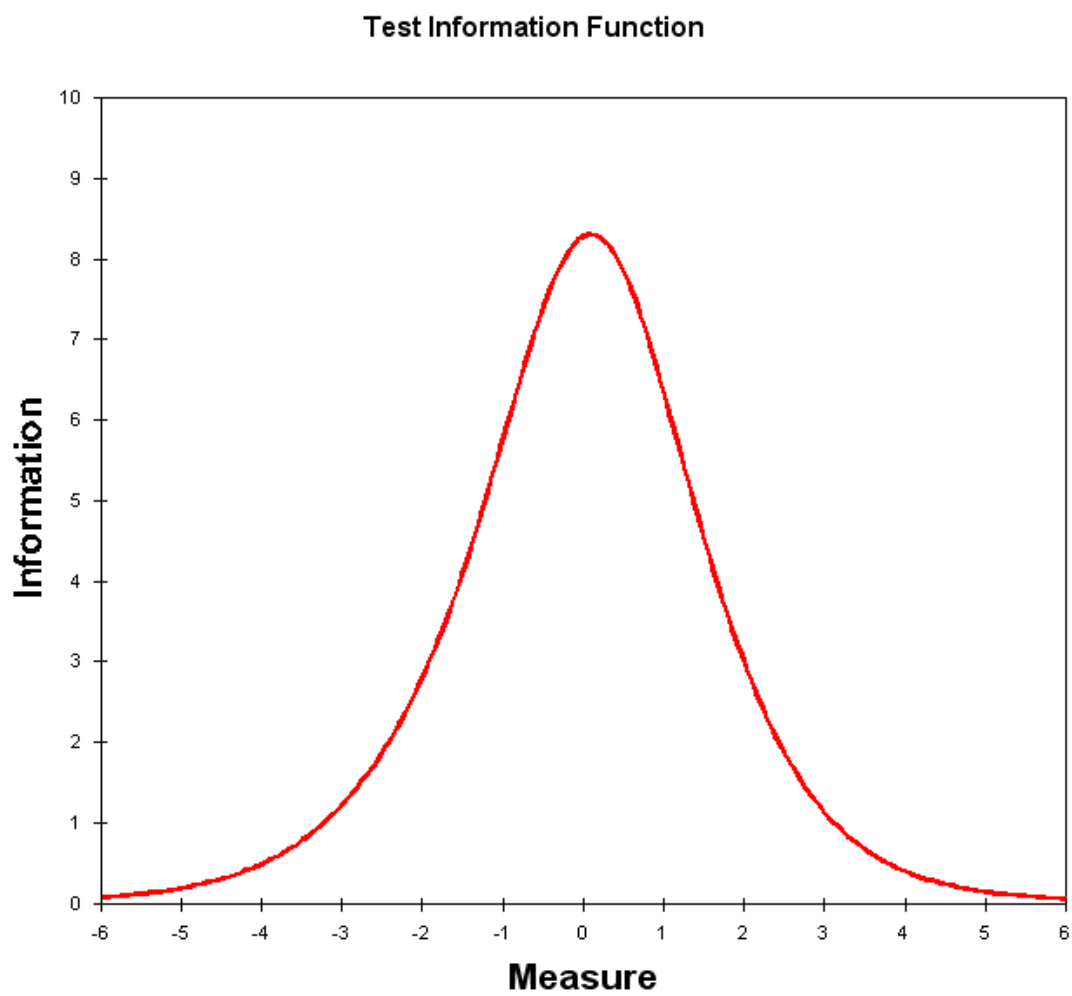


Figure D.6. Test Information Function – Grade 8 Mathematics

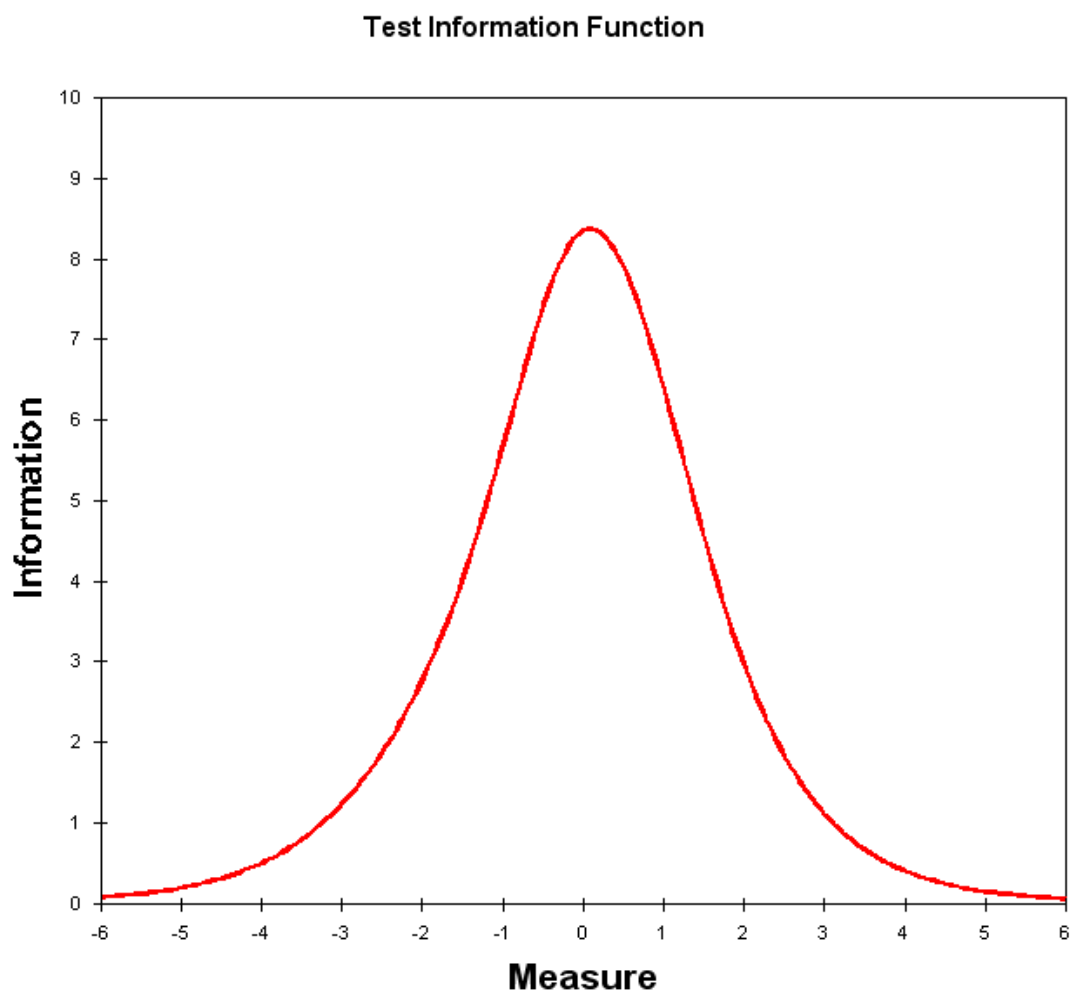


Figure D.7. Test Information Function – Grade 11 Mathematics

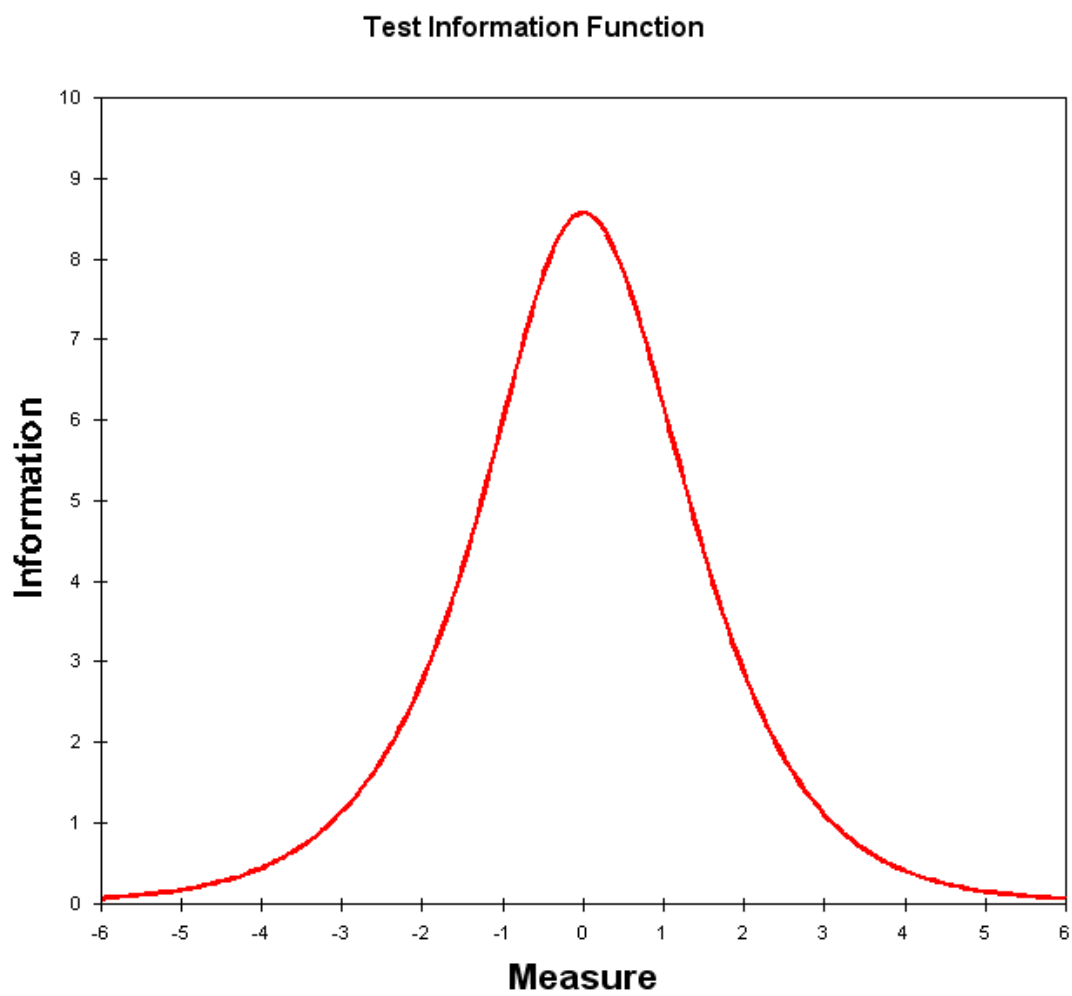


Figure D.8. Test Information Function – Grade 3 Reading

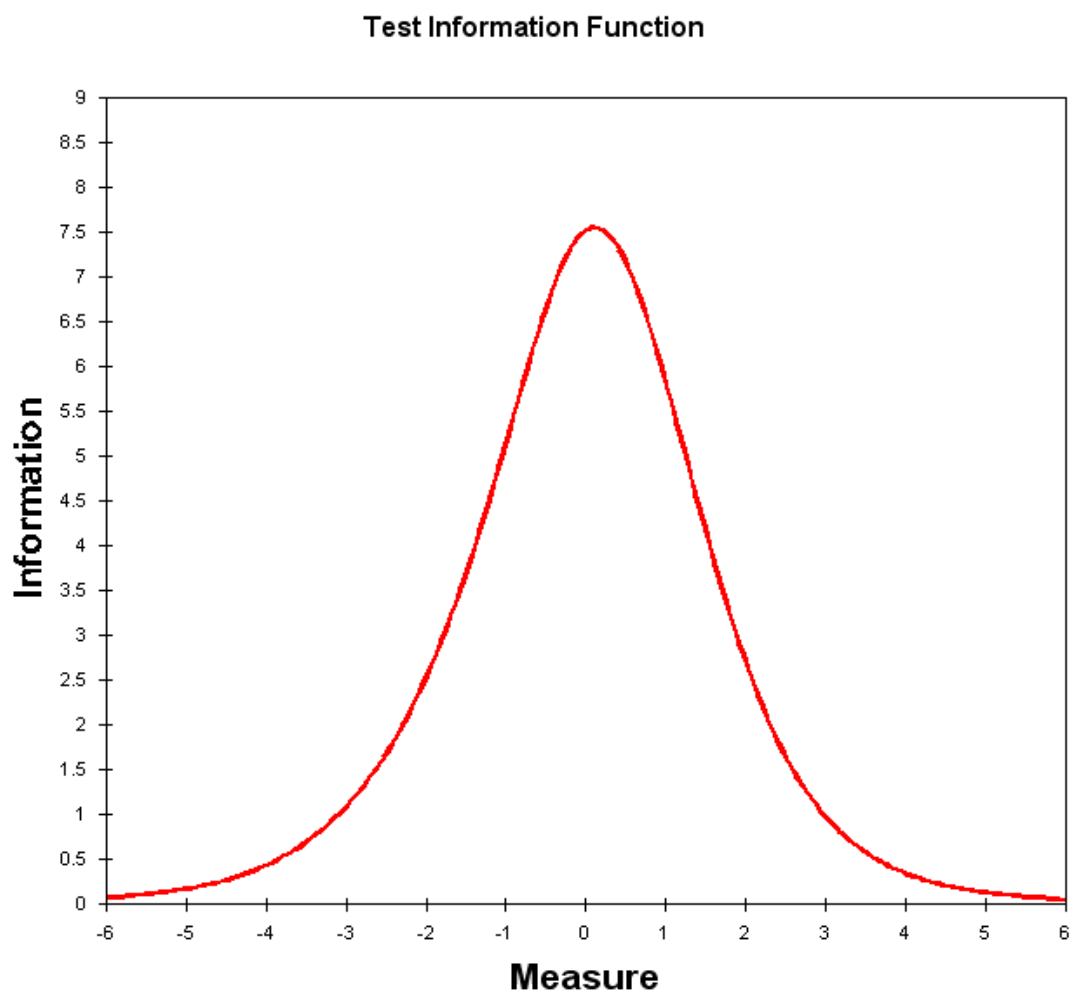


Figure D.9. Test Information Function – Grade 4 Reading

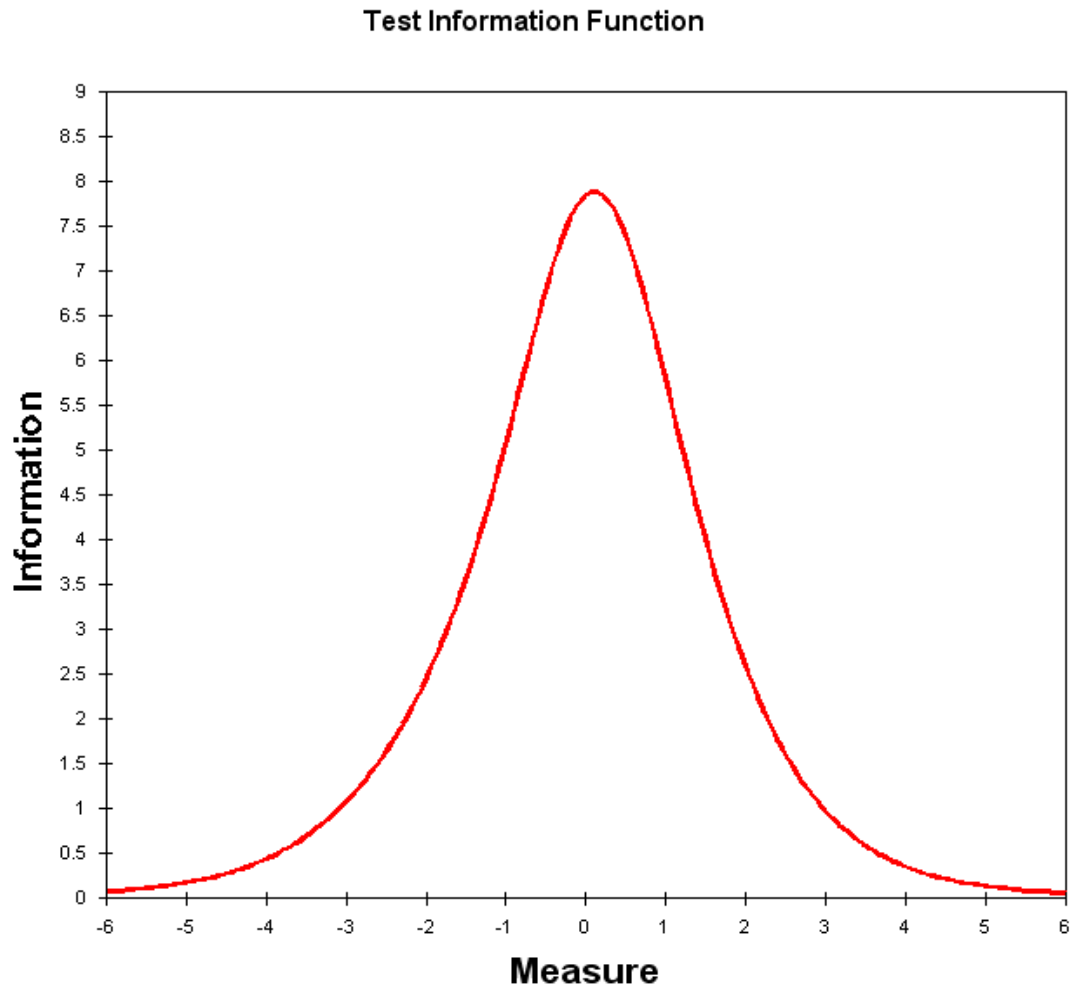


Figure D.10. Test Information Function – Grade 5 Reading

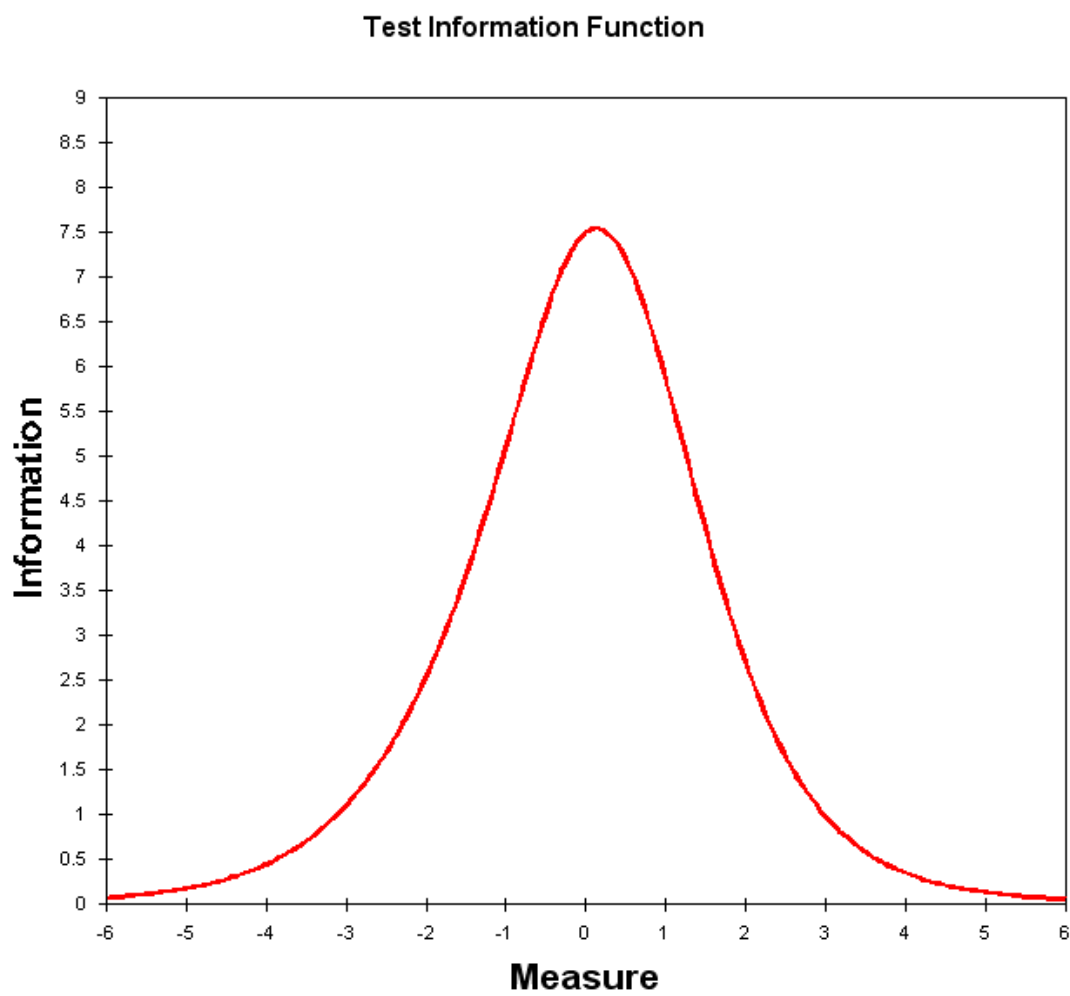


Figure D.11. Test Information Function – Grade 6 Reading

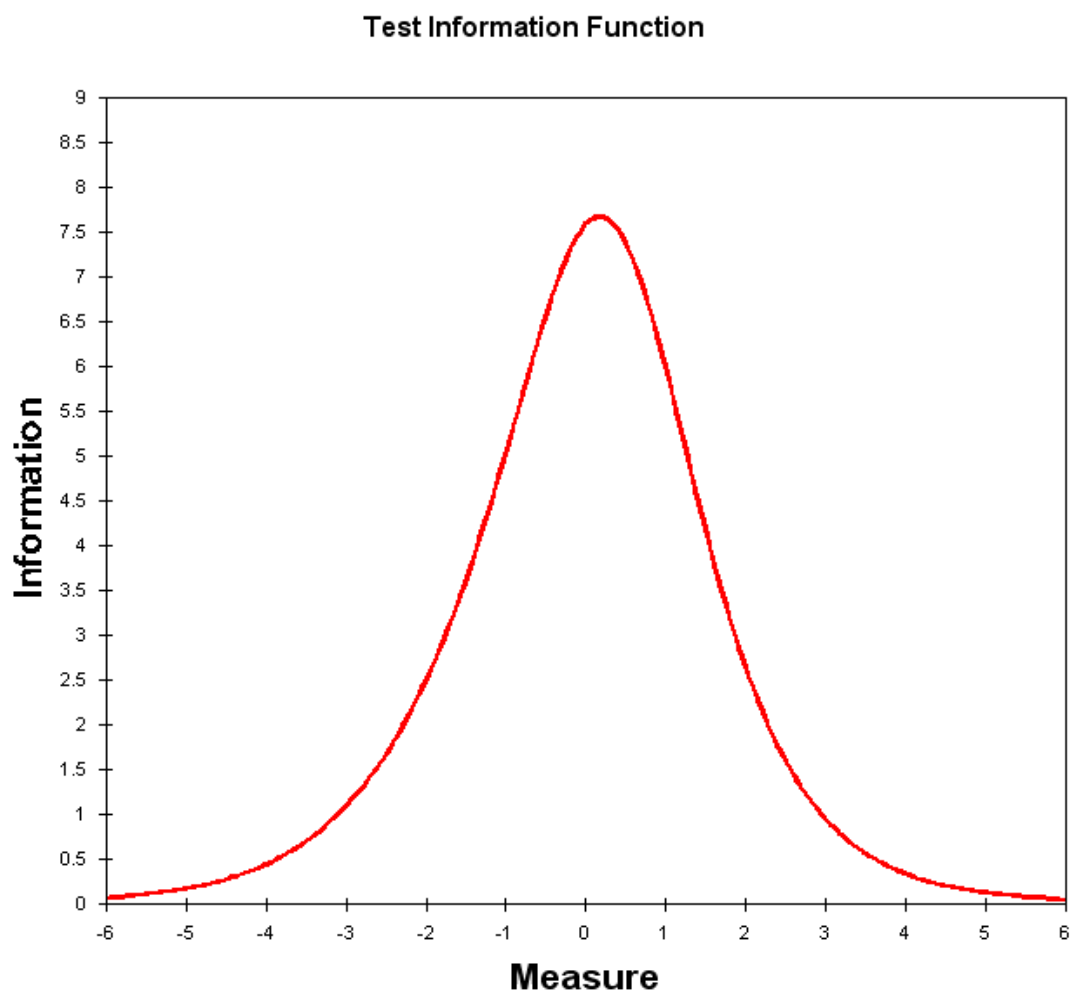


Figure D.12. Test Information Function – Grade 7 Reading

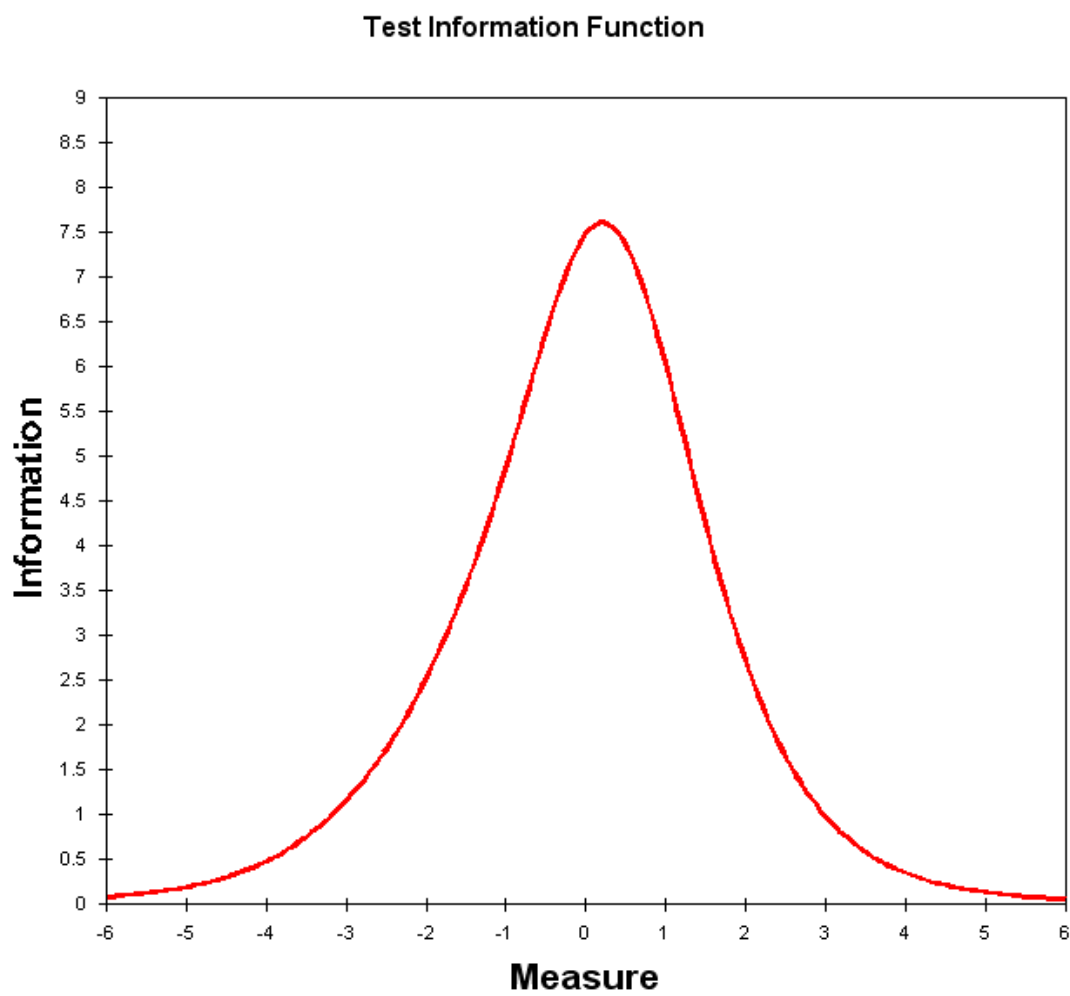


Figure D.13. Test Information Function – Grade 8 Reading

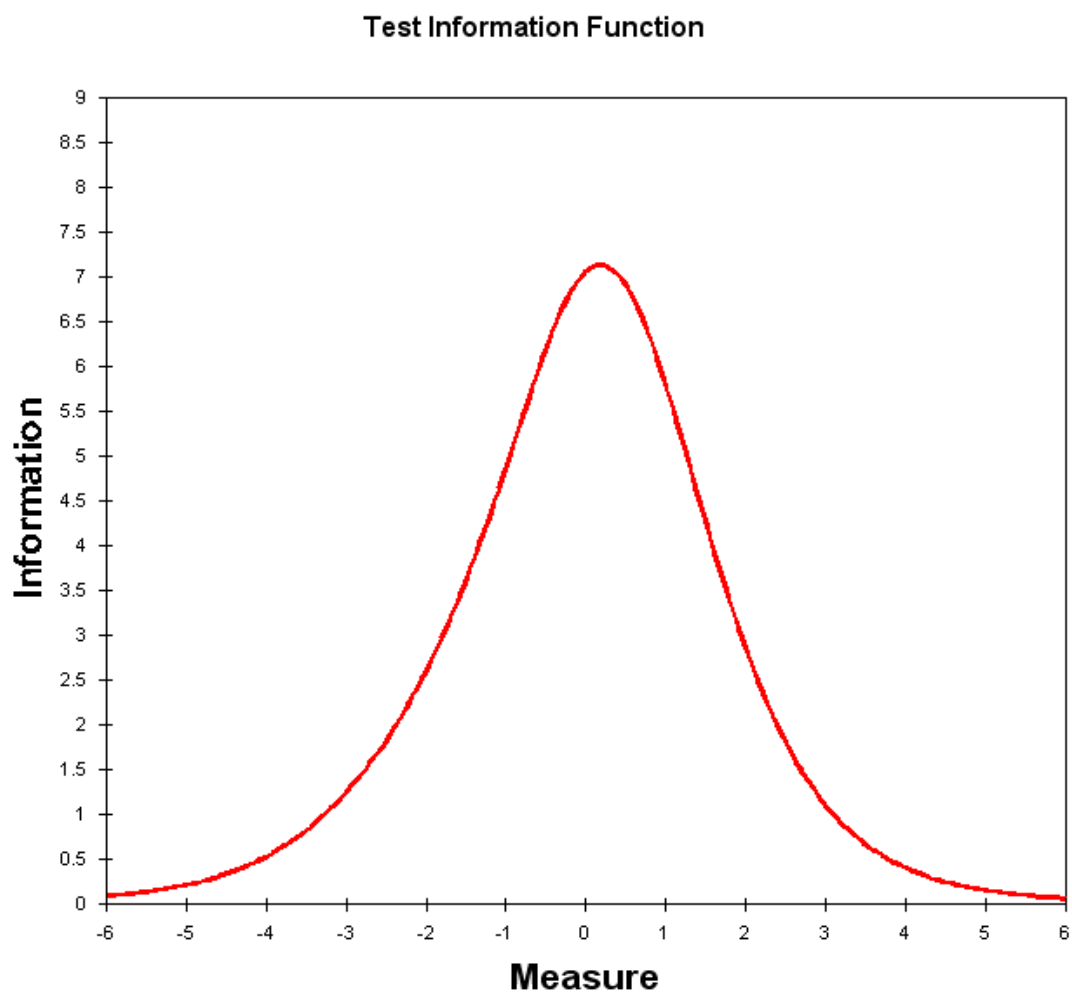


Figure D.14. Test Information Function – Grade 11 Reading

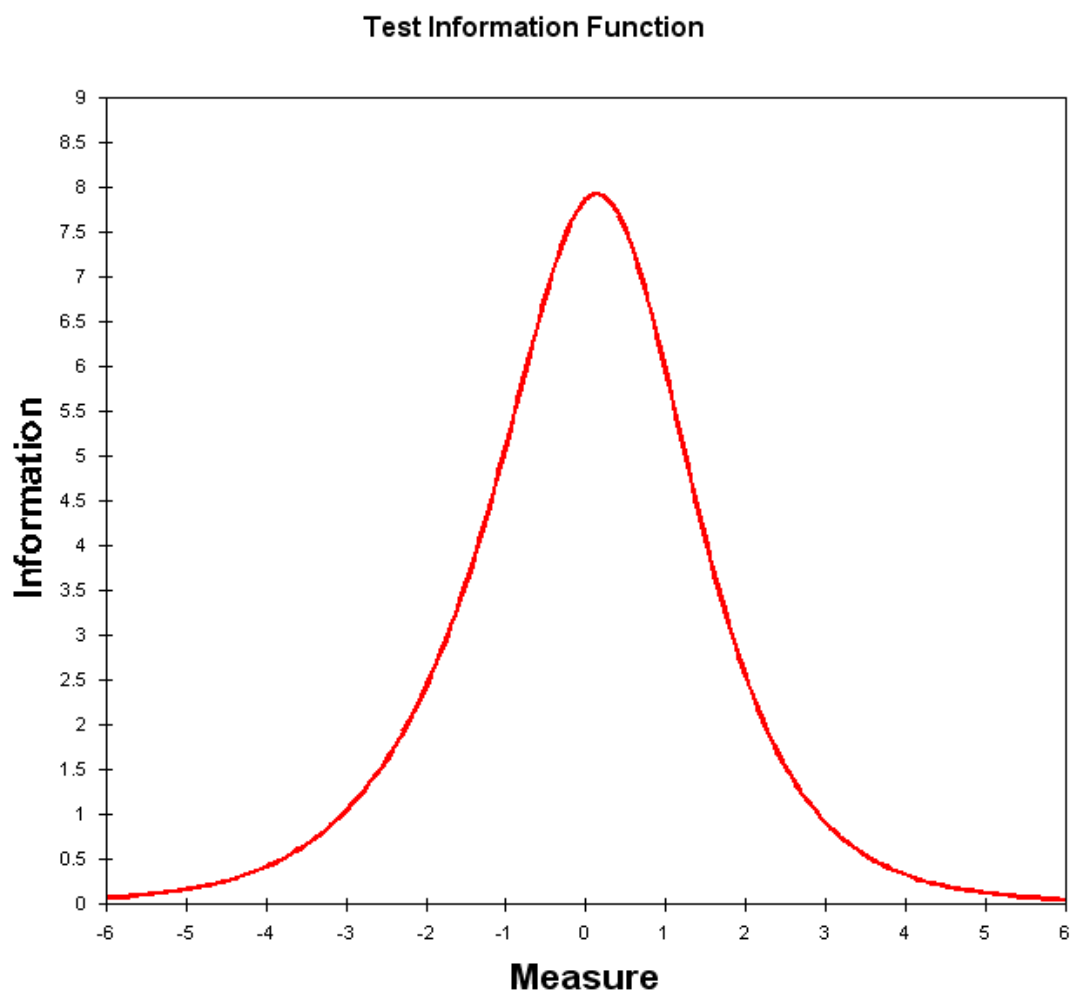


Figure D.15. Test Information Function – Grade 4 Science

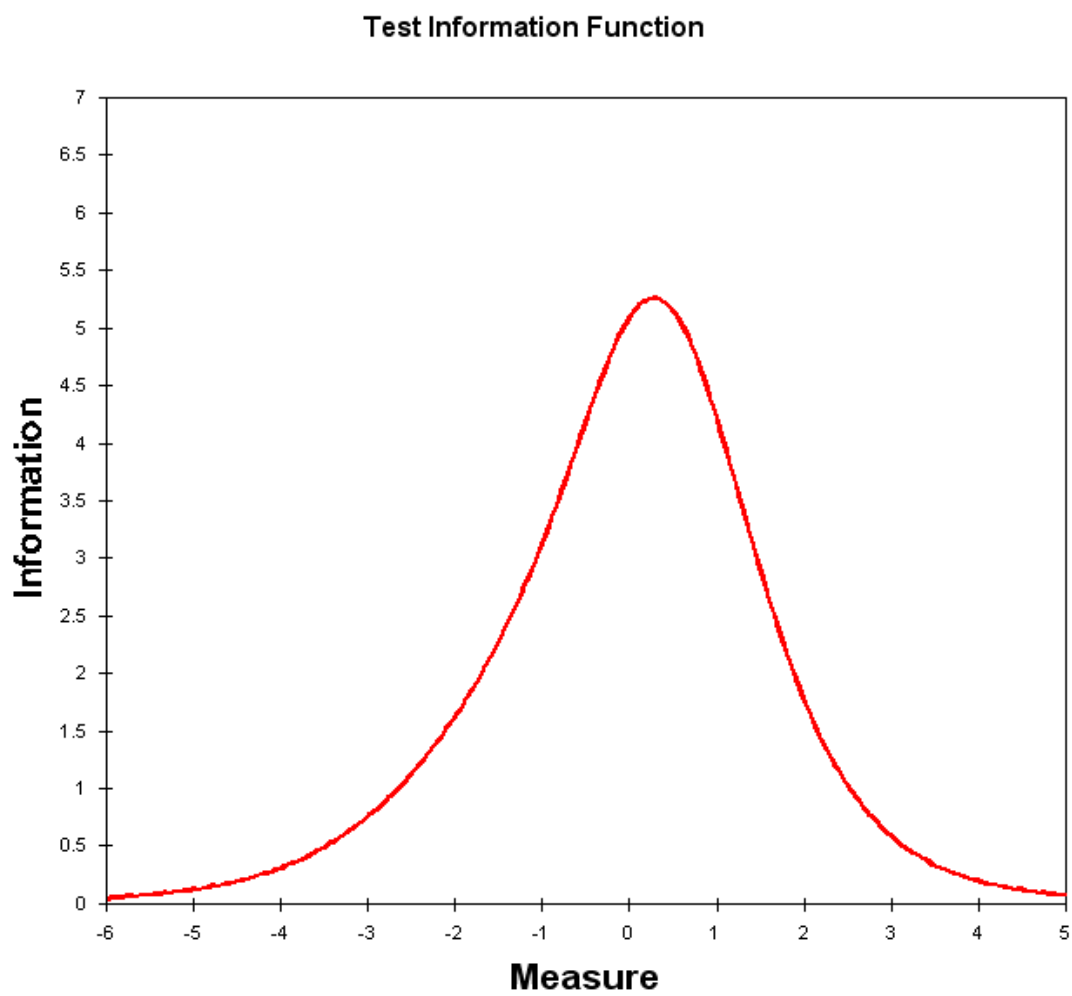


Figure D.16. Test Information Function – Grade 7 Science

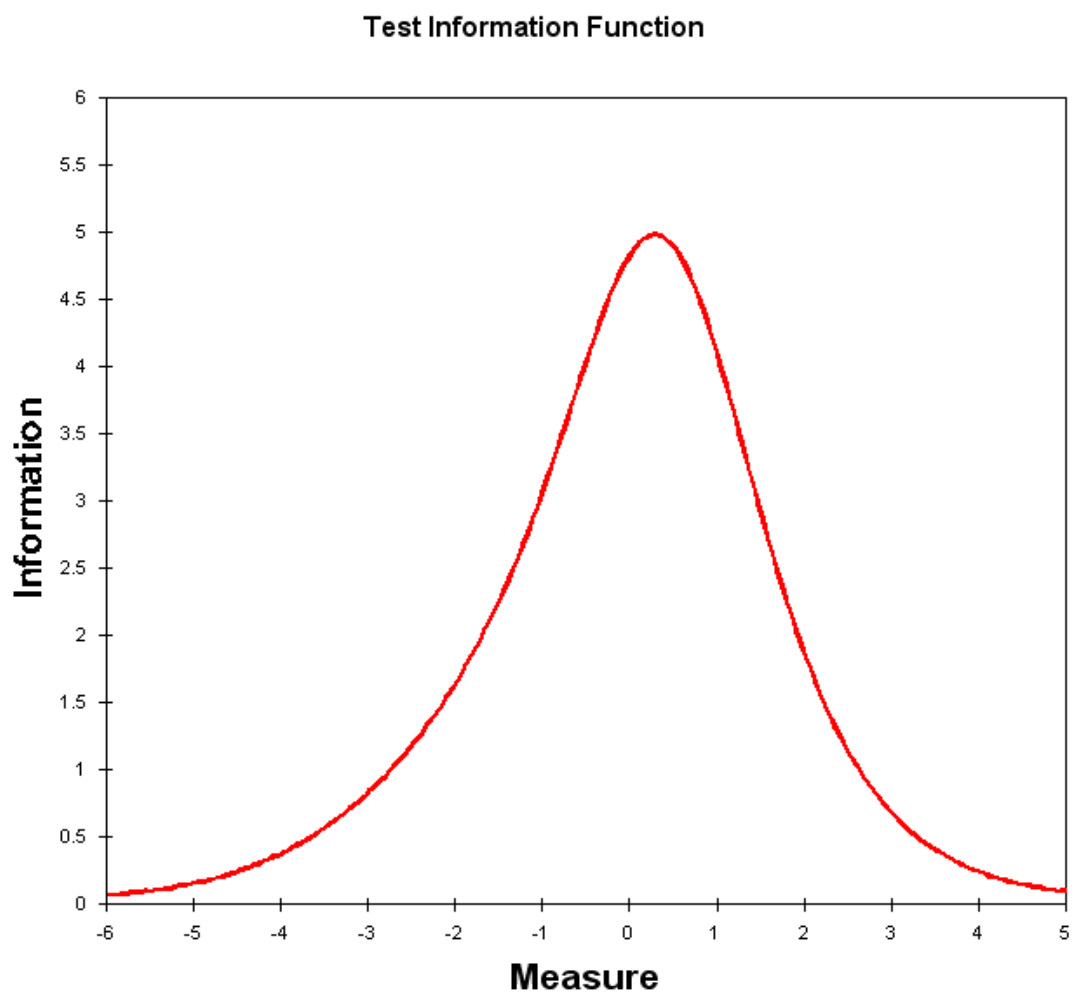


Figure D.17. Test Information Function – Grade 11 Science

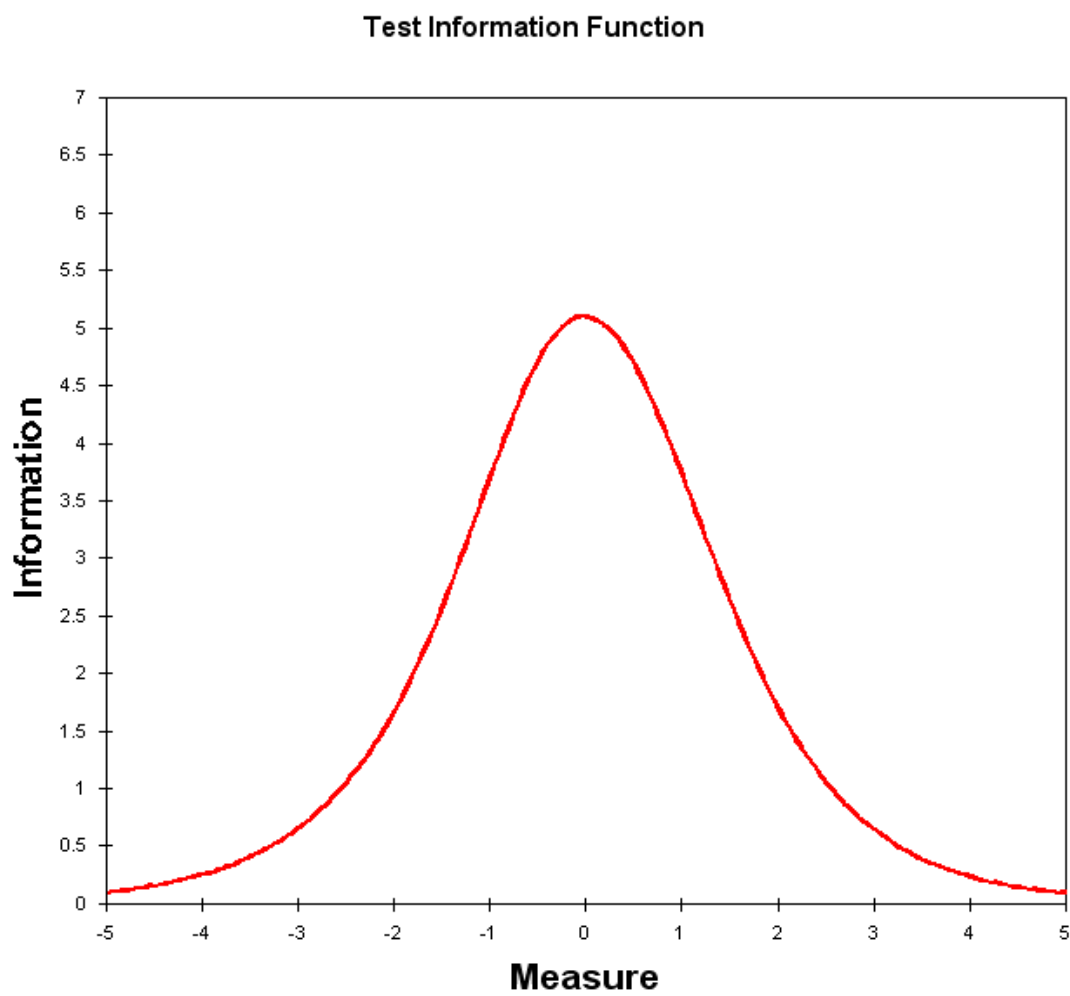


Figure D.18. Test Information Function – Grade 5 Writing

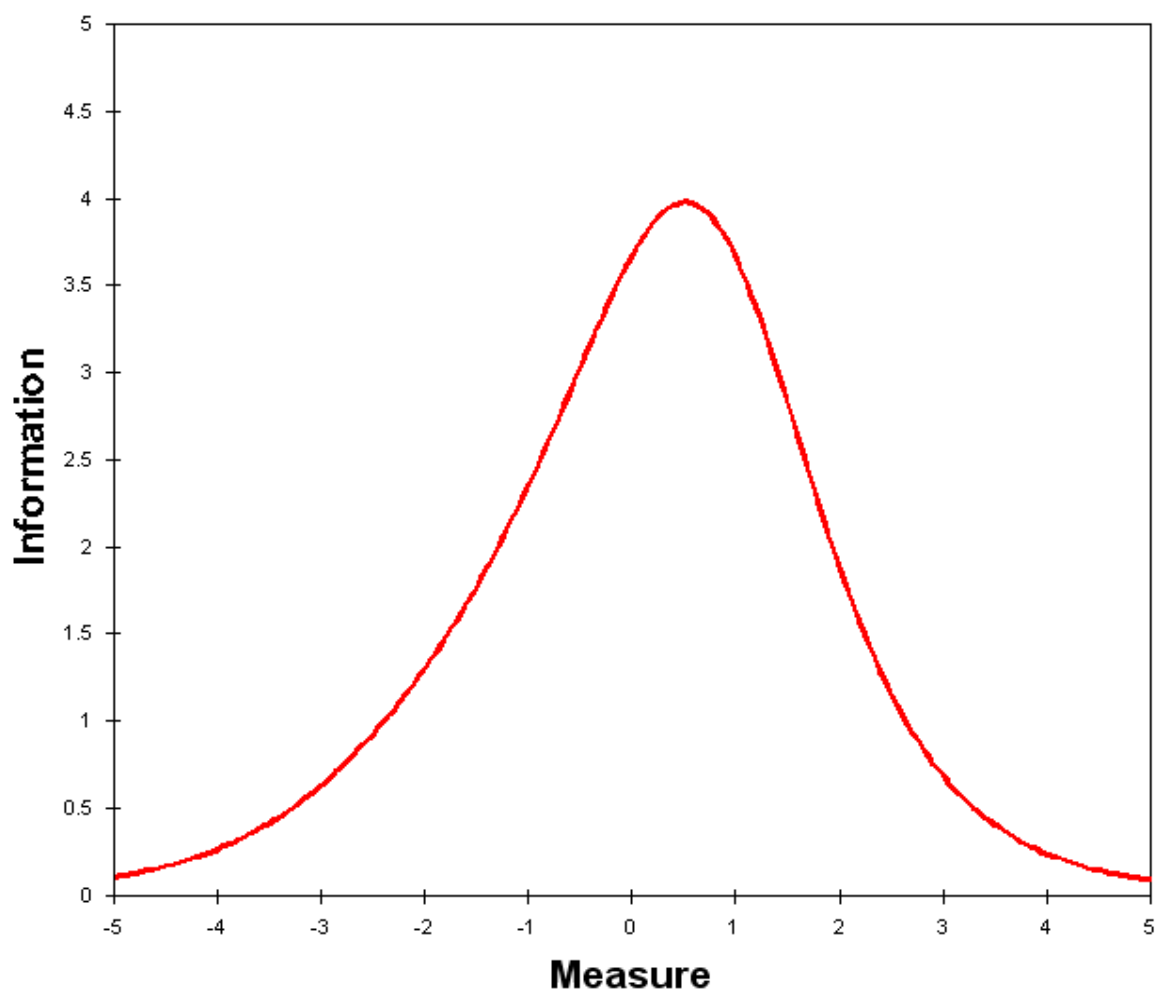


Figure D.19. Test Information Function – Grade 6 Writing

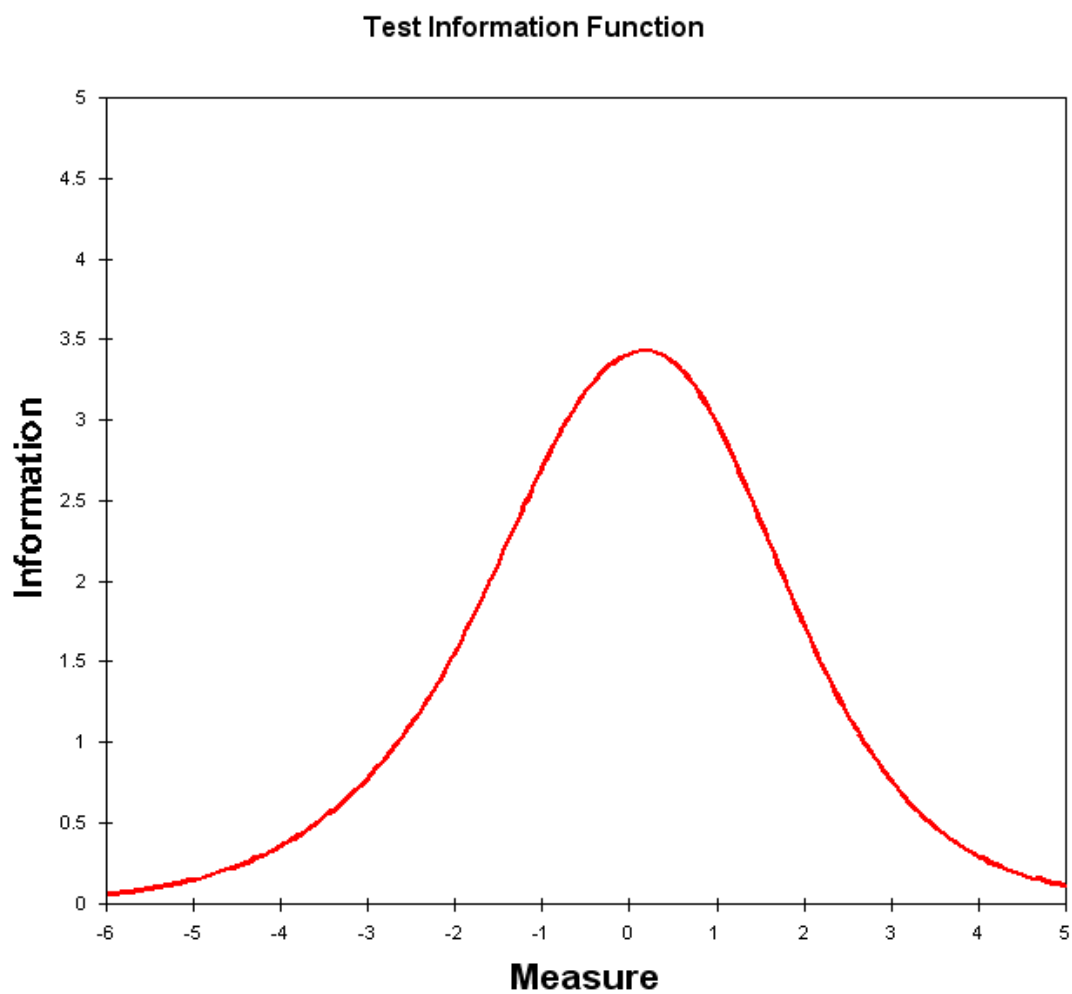


Figure D.20. Test Information Function – Grade 8 Writing

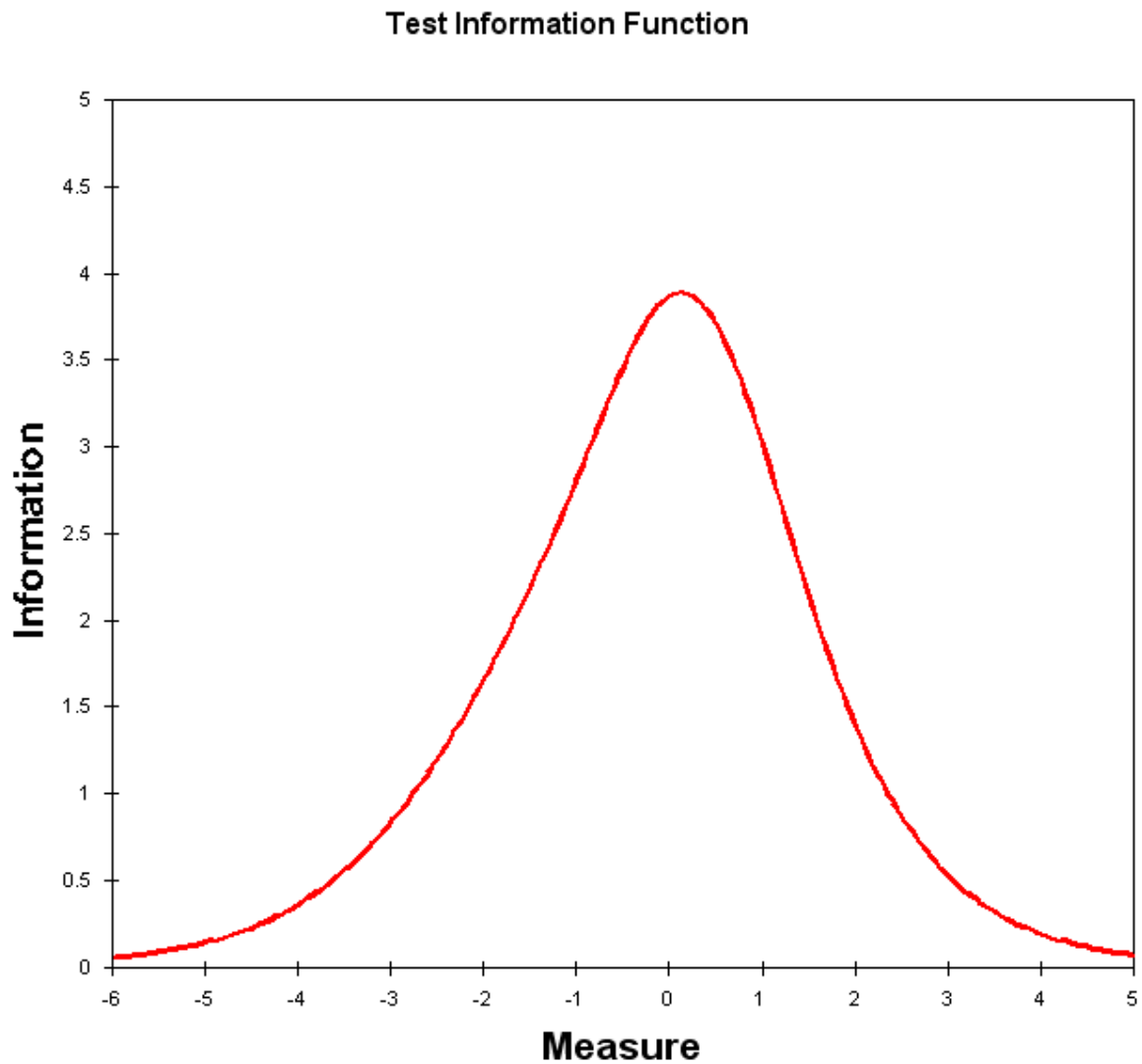
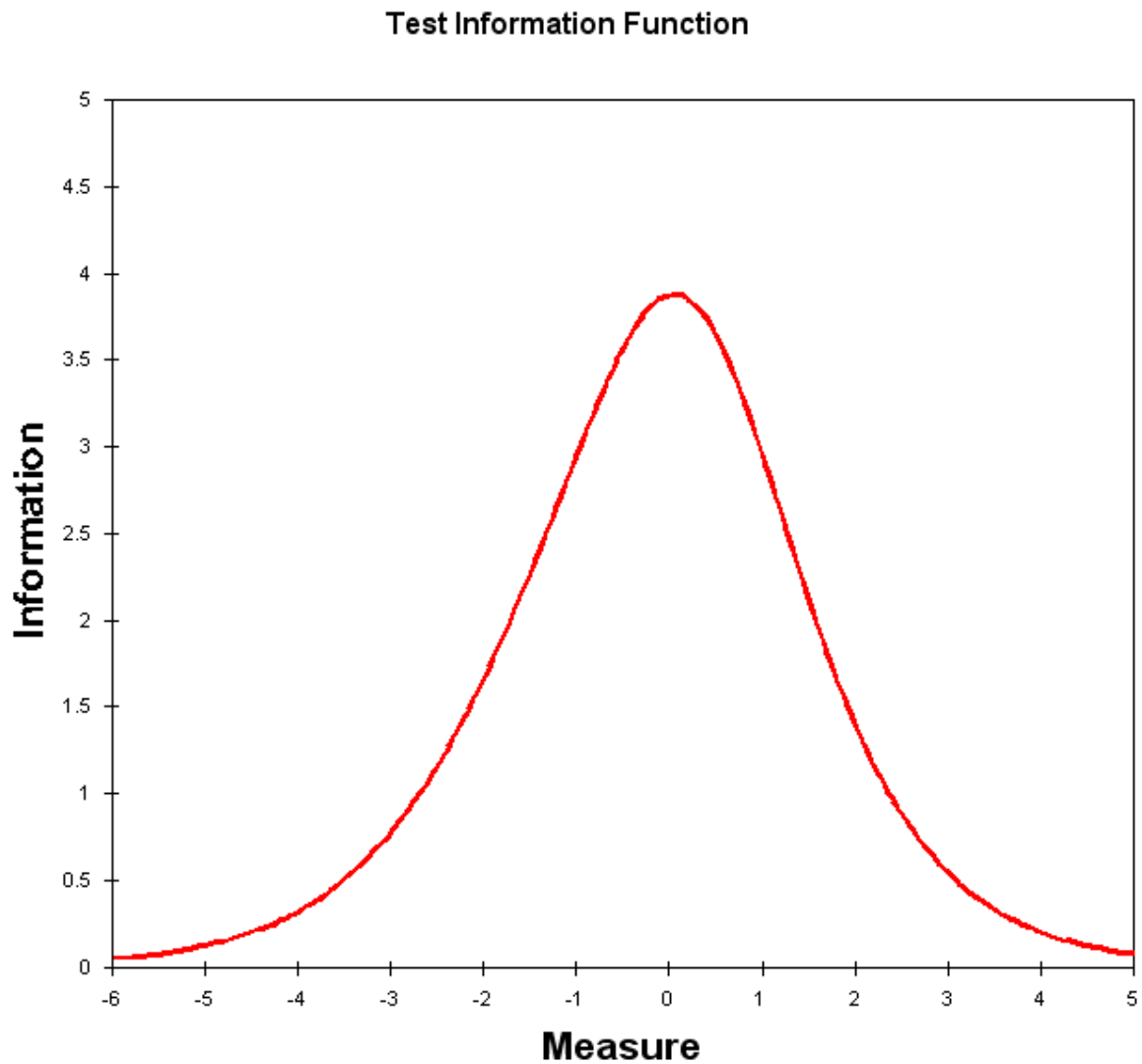


Figure D.21. Test Information Function – Grade 11 Writing



Appendix E
2008 IAA Standard Setting Item Maps

Grade 03 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	3M00 2	1	-2.89	6	A	1
2	3M04 9	1	-2.67	9	A	1
3	3M03 8	1	-2.35	8	A	1
4	3M03 6	1	-2.00	7	A,B,C	5
5	3M01 3	1	-1.98	6	A	7
6	3M06 1	1	-1.78	10	A,B	1
7	3M04 7	1	-1.71	8	C,D	4
8	3M05 6	1	-1.52	9	B	9
9	3M02 9	1	-1.46	7	A,B,C	1
10	3M02 2	1	-1.29	6	B,C	10
11	3M00 2	2	-0.83	6	A	1
12	3M04 9	2	-0.48	9	A	1
13	3M03 6	2	0.00	7	A,B,C	5
14	3M03 8	2	0.02	8	A	1
15	3M01 3	2	0.22	6	A	7
16	3M00 2	3	0.27	6	A	1

Grade 03 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	3M06 1	2	0.32	10	A,B	1
18	3M04 7	2	0.37	8	C,D	4
19	3M05 6	2	0.40	9	B	9
20	3M04 9	3	0.52	9	A	1
21	3M02 2	2	0.64	6	B,C	10
22	3M02 9	2	0.65	7	A,B,C	1
23	3M03 6	3	0.74	7	A,B,C	5
24	3M03 8	3	0.78	8	A	1
25	3M01 3	3	0.92	6	A	7
26	3M04 7	3	1.16	8	C,D	4
27	3M06 1	3	1.18	10	A,B	1
28	3M05 6	3	1.25	9	B	9
29	3M02 9	3	1.39	7	A,B,C	1
30	3M02 2	3	1.52	6	B,C	10
31	3M00 2	4	1.62	6	A	1
32	3M04 9	4	1.70	9	A	1
33	3M03 6	4	1.91	7	A,B,C	5

Grade 03 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	3M03 8	4	1.93	8	A	1
35	3M01 3	4	2.10	6	A	7
36	3M04 7	4	2.16	8	C,D	4
37	3M06 1	4	2.19	10	A,B	1
38	3M05 6	4	2.22	9	B	9
39	3M02 2	4	2.33	6	B,C	10
40	3M02 9	4	2.49	7	A,B,C	1

Grade 04 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	4M040	1	-2.21	8	A	1
2	4M002	1	-2.18	6	A	1
3	4M056	1	-2.13	9	B	12
4	4M036	1	-2.09	7	A,B,C	4
5	4M005	1	-2.08	6	A	2
6	4M049	1	-2.05	9	A	1
7	4M061	1	-1.67	10	A,B	1
8	4M031	1	-1.66	7	A,B,C	2
9	4M045	1	-1.13	8	C,D	7
10	4M024	1	-0.93	6	B,C	11
11	4M056	2	-0.18	9	B	12
12	4M002	2	-0.06	6	A	1
13	4M005	2	-0.03	6	A	2
14	4M036	2	-0.03	7	A,B,C	4
15	4M049	2	0.06	9	A	1
16	4M040	2	0.21	8	A	1

Grade 04 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	4M03 1	2	0.23	7	A,B,C	2
18	4M06 1	2	0.24	10	A,B	1
19	4M03 6	3	0.74	7	A,B,C	4
20	4M00 5	3	0.75	6	A	2
21	4M02 4	2	0.76	6	B,C	11
22	4M04 5	2	0.76	8	C,D	7
23	4M05 6	3	0.78	9	B	12
24	4M00 2	3	0.79	6	A	1
25	4M06 1	3	1.04	10	A,B	1
26	4M04 0	3	1.05	8	A	1
27	4M03 1	3	1.08	7	A,B,C	2
28	4M04 9	3	1.18	9	A	1
29	4M02 4	3	1.49	6	B,C	11
30	4M04 5	3	1.73	8	C,D	7
31	4M05 6	4	2.03	9	B	12
32	4M00 5	4	2.11	6	A	2
33	4M03 6	4	2.13	7	A,B,C	4

Grade 04 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	4M04 9	4	2.14	9	A	1
35	4M00 2	4	2.15	6	A	1
36	4M04 0	4	2.22	8	A	1
37	4M03 1	4	2.28	7	A,B,C	2
38	4M06 1	4	2.33	10	A,B	1
39	4M04 5	4	2.58	8	C,D	7
40	4M02 4	4	2.66	6	B,C	11

Grade 05 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	5M03 5	1	-2.25	7	A,B,C	4
2	5M05 1	1	-2.21	9	A	1
3	5M02 6	1	-2.02	7	A,B,C	2
4	5M03 9	1	-2.00	8	A	1
5	5M06 2	1	-1.91	10	A,B	1
6	5M05 6	1	-1.86	9	B	14
7	5M00 2	1	-1.56	6	A	1
8	5M01 6	1	-1.08	6	B,C	12
9	5M04 3	1	-1.07	8	C,D	8
10	5M01 2	1	-0.93	6	A	10
11	5M03 5	2	-0.28	7	A,B,C	4
12	5M03 9	2	-0.15	8	A	1
13	5M05 1	2	-0.09	9	A	1
14	5M05 6	2	-0.06	9	B	14
15	5M02 6	2	0.14	7	A,B,C	2
16	5M06 2	2	0.20	10	A,B	1

Grade 05 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	5M00 2	2	0.39	6	A	1
18	5M01 6	2	0.49	6	B,C	12
19	5M04 3	2	0.58	8	C,D	8
20	5M01 2	2	0.64	6	A	10
21	5M03 5	3	0.67	7	A,B,C	4
22	5M03 9	3	0.68	8	A	1
23	5M05 1	3	0.75	9	A	1
24	5M05 6	3	0.84	9	B	14
25	5M06 2	3	0.94	10	A,B	1
26	5M02 6	3	0.96	7	A,B,C	2
27	5M00 2	3	1.18	6	A	1
28	5M01 6	3	1.25	6	B,C	12
29	5M04 3	3	1.35	8	C,D	8
30	5M01 2	3	1.57	6	A	10
31	5M03 5	4	1.62	7	A,B,C	4
32	5M03 9	4	1.73	8	A	1
33	5M05 1	4	1.73	9	A	1

Grade 05 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	5M05 6	4	1.73	9	B	14
35	5M02 6	4	1.81	7	A,B,C	2
36	5M06 2	4	1.85	10	A,B	1
37	5M00 2	4	1.98	6	A	1
38	5M01 6	4	2.11	6	B,C	12
39	5M04 3	4	2.16	8	C,D	8
40	5M01 2	4	2.28	6	A	10

Grade 06 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	6M05 3	1	-2.69	9	B	12
2	6M02 6	1	-2.37	7	A,B,C	3
3	6M06 4	1	-2.23	10	C	6
4	6M04 3	1	-2.19	9	A	2
5	6M06 1	1	-2.17	10	A,B	1
6	6M02 8	1	-2.13	8	A	1
7	6M03 5	1	-1.93	8	B	6
8	6M00 5	1	-1.75	6	A	4
9	6M02 1	1	-1.75	7	A,B,C	1
10	6M01 5	1	-1.19	6	B,C	12
11	6M05 3	2	-0.23	9	B	12
12	6M02 6	2	-0.08	7	A,B,C	3
13	6M06 4	2	0.02	10	C	6
14	6M02 8	2	0.11	8	A	1
15	6M04 3	2	0.11	9	A	2
16	6M03 5	2	0.18	8	B	6

Grade 06 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	6M06 1	2	0.20	10	A,B	1
18	6M00 5	2	0.37	6	A	4
19	6M02 1	2	0.49	7	A,B,C	1
20	6M02 6	3	0.68	7	A,B,C	3
21	6M05 3	3	0.73	9	B	12
22	6M01 5	2	0.74	6	B,C	12
23	6M06 4	3	0.92	10	C	6
24	6M06 1	3	0.97	10	A,B	1
25	6M02 8	3	1.03	8	A	1
26	6M03 5	3	1.14	8	B	6
27	6M04 3	3	1.22	9	A	2
28	6M02 1	3	1.43	7	A,B,C	1
29	6M00 5	3	1.46	6	A	4
30	6M01 5	3	1.76	6	B,C	12
31	6M05 3	4	2.31	9	B	12
32	6M02 6	4	2.34	7	A,B,C	3
33	6M06 4	4	2.37	10	C	6

Grade 06 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	6M04 3	4	2.43	9	A	2
35	6M02 8	4	2.47	8	A	1
36	6M06 1	4	2.61	10	A,B	1
37	6M00 5	4	2.62	6	A	4
38	6M03 5	4	2.64	8	B	6
39	6M02 1	4	2.72	7	A,B,C	1
40	6M01 5	4	3.08	6	B,C	12

Grade 07 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	7M054	1	-2.60	9	B	14
2	7M024	1	-1.98	7	A,B,C	3
3	7M023	1	-1.85	7	A,B,C	1
4	7M030	1	-1.85	8	A	1
5	7M003	1	-1.79	6	A	3
6	7M043	1	-1.79	9	A	5
7	7M057	1	-1.55	10	A,B	1
8	7M066	1	-1.50	10	C	7
9	7M012	1	-1.39	6	D	16
10	7M039	1	-1.14	8	C,D	12
11	7M054	2	-0.41	9	B	14
12	7M024	2	-0.14	7	A,B,C	3
13	7M003	2	-0.10	6	A	3
14	7M023	2	0.01	7	A,B,C	1
15	7M030	2	0.14	8	A	1
16	7M043	2	0.19	9	A	5

Grade 07 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	7M06 6	2	0.21	10	C	7
18	7M05 7	2	0.32	10	A,B	1
19	7M05 4	3	0.37	9	B	14
20	7M01 2	2	0.44	6	D	16
21	7M03 9	2	0.61	8	C,D	12
22	7M02 4	3	0.66	7	A,B,C	3
23	7M02 3	3	0.74	7	A,B,C	1
24	7M03 0	3	0.83	8	A	1
25	7M04 3	3	0.91	9	A	5
26	7M00 3	3	0.93	6	A	3
27	7M06 6	3	0.95	10	C	7
28	7M05 7	3	1.13	10	A,B	1
29	7M01 2	3	1.23	6	D	16
30	7M03 9	3	1.32	8	C,D	12
31	7M05 4	4	1.33	9	B	14
32	7M02 4	4	1.53	7	A,B,C	3
33	7M00 3	4	1.57	6	A	3

Grade 07 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	7M023	4	1.65	7	A,B,C	1
35	7M030	4	1.66	8	A	1
36	7M057	4	1.72	10	A,B	1
37	7M043	4	1.74	9	A	5
38	7M066	4	1.75	10	C	7
39	7M012	4	1.90	6	D	16
40	7M039	4	2.00	8	C,D	12

Grade 08 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	8M025	1	-2.48	7	A,B,C	3
2	8M021	1	-2.25	7	A,B,C	1
3	8M049	1	-2.21	9	B	10
4	8M053	1	-1.93	10	A,B	1
5	8M001	1	-1.92	6	A	3
6	8M037	1	-1.57	8	C,D	13
7	8M058	1	-1.52	10	A,B	3
8	8M029	1	-1.48	8	A	1
9	8M010	1	-1.37	6	B,C	9
10	8M045	1	-1.23	9	A	5
11	8M025	2	-0.35	7	A,B,C	3
12	8M049	2	-0.30	9	B	10
13	8M021	2	-0.28	7	A,B,C	1
14	8M001	2	-0.04	6	A	3
15	8M053	2	0.06	10	A,B	1
16	8M058	2	0.24	10	A,B	3

Grade 08 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	8M010	2	0.40	6	B,C	9
18	8M037	2	0.41	8	C,D	13
19	8M029	2	0.46	8	A	1
20	8M025	3	0.67	7	A,B,C	3
21	8M045	2	0.67	9	A	5
22	8M049	3	0.70	9	B	10
23	8M021	3	0.72	7	A,B,C	1
24	8M053	3	0.90	10	A,B	1
25	8M001	3	1.03	6	A	3
26	8M058	3	1.24	10	A,B	3
27	8M037	3	1.36	8	C,D	13
28	8M010	3	1.44	6	B,C	9
29	8M029	3	1.50	8	A	1
30	8M045	3	1.80	9	A	5
31	8M025	4	2.03	7	A,B,C	3
32	8M049	4	2.07	9	B	10
33	8M021	4	2.10	7	A,B,C	1

Grade 08 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	8M001	4	2.13	6	A	3
35	8M053	4	2.25	10	A,B	1
36	8M058	4	2.40	10	A,B	3
37	8M029	4	2.45	8	A	1
38	8M037	4	2.46	8	C,D	13
39	8M010	4	2.48	6	B,C	9
40	8M045	4	2.56	9	A	5

Grade 11 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	11M04 9	1	-2.58	9	A	6
2	11M06 1	1	-2.43	10	A,B	1
3	11M05 9	1	-2.20	9	A	7
4	11M04 2	1	-2.18	8	B	12
5	11M03 7	1	-2.11	8	A	4
6	11M00 2	1	-2.04	6	A	1
7	11M01 7	1	-1.84	6	B,C	13
8	11M03 3	1	-1.59	7	A,B,C	3
9	11M02 7	1	-1.50	7	A,B,C	1
10	11M02 4	1	-1.38	6	D	18
11	11M04 9	2	-0.24	9	A	6
12	11M06 1	2	-0.07	10	A,B	1
13	11M05 9	2	0.10	9	A	7
14	11M00 2	2	0.15	6	A	1
15	11M03 7	2	0.22	8	A	4
16	11M04 2	2	0.27	8	B	12

Grade 11 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	11M01 7	2	0.36	6	B,C	13
18	11M02 7	2	0.43	7	A,B,C	1
19	11M03 3	2	0.45	7	A,B,C	3
20	11M02 4	2	0.59	6	D	18
21	11M04 9	3	0.87	9	A	6
22	11M06 1	3	0.89	10	A,B	1
23	11M04 2	3	0.92	8	B	12
24	11M00 2	3	0.98	6	A	1
25	11M03 7	3	0.98	8	A	4
26	11M05 9	3	0.98	9	A	7
27	11M01 7	3	1.19	6	B,C	13
28	11M03 3	3	1.20	7	A,B,C	3
29	11M02 7	3	1.36	7	A,B,C	1
30	11M04 9	4	1.56	9	A	6
31	11M02 4	3	1.62	6	D	18
32	11M05 9	4	1.65	9	A	7
33	11M06 1	4	1.65	10	A,B	1

Grade 11 Math Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	11M03 7	4	1.78	8	A	4
35	11M00 2	4	1.79	6	A	1
36	11M04 2	4	1.83	8	B	12
37	11M01 7	4	1.92	6	B,C	13
38	11M02 4	4	1.98	6	D	18
39	11M03 3	4	2.02	7	A,B,C	3
40	11M02 7	4	2.04	7	A,B,C	1

Grade 03 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	3R008	1	-2.62	01	A	008
2	3R011	1	-2.45	01	A	011
3	3R044	1	-2.42	02	A	007
4	3R029	1	-2.32	01	C	020
5	3R022	1	-2.28	01	B,C	014
6	3R001	1	-2.25	01	A	003
7	3R018	1	-2.16	01	B,C	013
8	3R046	1	-1.93	02	A	007
9	3R051	1	-1.36	02	B	010
10	3R011	2	-0.35	01	A	011
11	3R008	2	-0.19	01	A	008
12	3R029	2	-0.11	01	C	020
13	3R044	2	0.01	02	A	007
14	3R022	2	0.07	01	B,C	014
15	3R046	2	0.18	02	A	007
16	3R001	2	0.24	01	A	003

Grade 03 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	3R018	2	0.27	01	B,C	013
18	3R044	3	0.76	02	A	007
19	3R051	2	0.81	02	B	010
20	3R011	3	0.82	01	A	011
21	3R008	3	0.96	01	A	008
22	3R022	3	1.22	01	B,C	014
23	3R001	3	1.23	01	A	003
24	3R029	3	1.25	01	C	020
25	3R018	3	1.31	01	B,C	013
26	3R046	3	1.41	02	A	007
27	3R051	3	1.68	02	B	010
28	3R011	4	2.33	01	A	011
29	3R044	4	2.38	02	A	007
30	3R008	4	2.39	01	A	008
31	3R029	4	2.49	01	C	020
32	3R018	4	2.51	01	B,C	013
33	3R046	4	2.52	02	A	007

Grade 03 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	3R00 1	4	2.55	01	A	003
35	3R02 2	4	2.56	01	B,C	014
36	3R05 1	4	2.78	02	B	010

Grade 04 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	4R008	1	-2.63	01	A	007
2	4R011	1	-2.28	01	B,C	009
3	4R049	1	-2.26	02	A	006
4	4R004	1	-2.22	01	A	004
5	4R018	1	-1.94	01	B,C	010
6	4R040	1	-1.91	02	A	001
7	4R035	1	-1.85	01	C	025
8	4R022	1	-1.80	01	B,C	015
9	4R054	1	-1.05	02	B	013
10	4R008	2	-0.26	01	A	007
11	4R011	2	-0.13	01	B,C	009
12	4R004	2	-0.12	01	A	004
13	4R049	2	-0.04	02	A	006
14	4R018	2	0.22	01	B,C	010
15	4R022	2	0.29	01	B,C	015
16	4R035	2	0.38	01	C	025

Grade 04 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	4R040	2	0.44	02	A	001
18	4R008	3	0.88	01	A	007
19	4R011	3	0.89	01	B,C	009
20	4R004	3	0.93	01	A	004
21	4R049	3	0.93	02	A	006
22	4R054	2	1.03	02	B	013
23	4R022	3	1.24	01	B,C	015
24	4R018	3	1.31	01	B,C	010
25	4R035	3	1.31	01	C	025
26	4R040	3	1.46	02	A	001
27	4R054	3	1.90	02	B	013
28	4R004	4	2.14	01	A	004
29	4R008	4	2.15	01	A	007
30	4R011	4	2.15	01	B,C	009
31	4R049	4	2.15	02	A	006
32	4R018	4	2.19	01	B,C	010
33	4R022	4	2.27	01	B,C	015

Grade 04 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	4R03 5	4	2.31	01	C	025
35	4R04 0	4	2.35	02	A	001
36	4R05 4	4	2.54	02	B	013

Grade 05 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	5R02 5	1	-2.47	01	C	016
2	5R01 2	1	-2.25	01	B,C	007
3	5R04 3	1	-2.07	02	A	001
4	5R00 2	1	-2.02	01	A	002
5	5R01 3	1	-1.98	01	B,C	008
6	5R05 6	1	-1.94	02	A	008
7	5R02 2	1	-1.71	01	B,C	013
8	5R03 2	1	-1.71	01	C	020
9	5R06 2	1	-1.28	02	B	014
10	5R02 5	2	-0.32	01	C	016
11	5R00 2	2	-0.10	01	A	002
12	5R01 2	2	0.02	01	B,C	007
13	5R05 6	2	0.03	02	A	008
14	5R04 3	2	0.08	02	A	001
15	5R01 3	2	0.18	01	B,C	008
16	5R02 2	2	0.41	01	B,C	013

Grade 05 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	5R03 2	2	0.46	01	C	020
18	5R06 2	2	0.66	02	B	014
19	5R02 5	3	0.74	01	C	016
20	5R00 2	3	1.02	01	A	002
21	5R01 2	3	1.08	01	B,C	007
22	5R05 6	3	1.19	02	A	008
23	5R04 3	3	1.23	02	A	001
24	5R01 3	3	1.25	01	B,C	008
25	5R02 2	3	1.37	01	B,C	013
26	5R03 2	3	1.54	01	C	020
27	5R06 2	3	1.59	02	B	014
28	5R02 5	4	2.12	01	C	016
29	5R00 2	4	2.20	01	A	002
30	5R01 2	4	2.20	01	B,C	007
31	5R05 6	4	2.24	02	A	008
32	5R04 3	4	2.26	02	A	001
33	5R01 3	4	2.31	01	B,C	008

Grade 05 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	5R03 2	4	2.37	01	C	020
35	5R02 2	4	2.39	01	B,C	013
36	5R06 2	4	2.49	02	B	014

Grade 06 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	6R050	1	-2.42	02	A	007
2	6R003	1	-2.28	01	A	003
3	6R008	1	-2.17	01	A	005
4	6R047	1	-2.10	02	A	001
5	6R021	1	-1.78	01	B	012
6	6R063	1	-1.58	02	A	014
7	6R040	1	-1.56	01	C	022
8	6R015	1	-1.53	01	B	011
9	6R037	1	-1.32	01	C	020
10	6R003	2	-0.22	01	A	003
11	6R047	2	-0.09	02	A	001
12	6R050	2	-0.09	02	A	007
13	6R008	2	0.18	01	A	005
14	6R021	2	0.27	01	B	012
15	6R040	2	0.34	01	C	022
16	6R015	2	0.35	01	B	011

Grade 06 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	6R06 3	2	0.48	02	A	014
18	6R03 7	2	0.65	01	C	020
19	6R04 7	3	0.82	02	A	001
20	6R05 0	3	0.86	02	A	007
21	6R00 3	3	0.89	01	A	003
22	6R00 8	3	1.02	01	A	005
23	6R01 5	3	1.15	01	B	011
24	6R02 1	3	1.26	01	B	012
25	6R04 0	3	1.27	01	C	022
26	6R06 3	3	1.35	02	A	014
27	6R03 7	3	1.60	01	C	020
28	6R00 3	4	2.03	01	A	003
29	6R05 0	4	2.05	02	A	007
30	6R04 7	4	2.11	02	A	001
31	6R00 8	4	2.16	01	A	005
32	6R04 0	4	2.25	01	C	022
33	6R06 3	4	2.25	02	A	014

Grade 06 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	6R02 1	4	2.28	01	B	012
35	6R01 5	4	2.31	01	B	011
36	6R03 7	4	2.35	01	C	020

Grade 07 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	7R02 1	1	-2.20	01	C	015
2	7R00 2	1	-2.18	01	A	003
3	7R01 1	1	-2.04	01	B	009
4	7R04 6	1	-2.02	02	A	001
5	7R00 5	1	-1.95	01	A	005
6	7R02 9	1	-1.76	01	C	017
7	7R05 4	1	-1.76	02	A	006
8	7R04 3	1	-1.56	01	C	022
9	7R05 7	1	-1.24	02	A	008
10	7R00 2	2	-0.26	01	A	003
11	7R02 1	2	-0.20	01	C	015
12	7R00 5	2	-0.12	01	A	005
13	7R04 6	2	-0.06	02	A	001
14	7R01 1	2	0.09	01	B	009
15	7R02 9	2	0.09	01	C	017
16	7R05 4	2	0.27	02	A	006

Grade 07 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	7R04 3	2	0.54	01	C	022
18	7R05 7	2	0.71	02	A	008
19	7R02 1	3	0.91	01	C	015
20	7R00 5	3	0.99	01	A	005
21	7R00 2	3	1.01	01	A	003
22	7R01 1	3	1.04	01	B	009
23	7R04 6	3	1.05	02	A	001
24	7R02 9	3	1.06	01	C	017
25	7R04 3	3	1.26	01	C	022
26	7R05 4	3	1.30	02	A	006
27	7R05 7	3	1.65	02	A	008
28	7R02 1	4	2.29	01	C	015
29	7R00 2	4	2.31	01	A	003
30	7R00 5	4	2.37	01	A	005
31	7R01 1	4	2.41	01	B	009
32	7R04 6	4	2.41	02	A	001
33	7R02 9	4	2.46	01	C	017

Grade 07 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	7R05 4	4	2.55	02	A	006
35	7R04 3	4	2.63	01	C	022
36	7R05 7	4	2.77	02	A	008

Grade 08 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	8R00 1	1	-2.67	01	A	003
2	8R03 2	1	-2.30	01	C	016
3	8R03 9	1	-2.12	01	C	023
4	8R00 8	1	-1.95	01	B	008
5	8R04 6	1	-1.87	02	A	004
6	8R04 4	1	-1.78	02	A	001
7	8R05 1	1	-1.76	02	A	006
8	8R02 2	1	-1.50	01	B	012
9	8R01 9	1	-1.46	01	B	010
10	8R00 1	2	-0.48	01	A	003
11	8R03 2	2	-0.22	01	C	016
12	8R05 1	2	0.13	02	A	006
13	8R04 4	2	0.14	02	A	001
14	8R00 8	2	0.17	01	B	008
15	8R03 9	2	0.17	01	C	023
16	8R04 6	2	0.26	02	A	004

Grade 08 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	8R02 2	2	0.48	01	B	012
18	8R01 9	2	0.60	01	B	010
19	8R00 1	3	0.75	01	A	003
20	8R03 2	3	0.94	01	C	016
21	8R00 8	3	1.14	01	B	008
22	8R04 4	3	1.32	02	A	001
23	8R05 1	3	1.35	02	A	006
24	8R03 9	3	1.38	01	C	023
25	8R04 6	3	1.44	02	A	004
26	8R02 2	3	1.62	01	B	012
27	8R01 9	3	1.73	01	B	010
28	8R00 1	4	2.36	01	A	003
29	8R03 2	4	2.43	01	C	016
30	8R00 8	4	2.57	01	B	008
31	8R03 9	4	2.63	01	C	023
32	8R04 4	4	2.68	02	A	001
33	8R04 6	4	2.68	02	A	004

Grade 08 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	8R05 1	4	2.68	02	A	006
35	8R02 2	4	2.82	01	B	012
36	8R01 9	4	2.94	01	B	010

Grade 11 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	11R003	1	-3.04	01	A	001
2	11R020	1	-2.78	01	C	009
3	11R014	1	-2.56	01	C	008
4	11R024	1	-2.42	01	C	013
5	11R051	1	-2.28	01	C	025
6	11R035	1	-2.21	01	C	014
7	11R007	1	-2.15	01	A	002
8	11R038	1	-1.90	01	C	016
9	11R044	1	-1.84	01	C	022
10	11R003	2	-0.18	01	A	001
11	11R020	2	-0.01	01	C	009
12	11R014	2	0.01	01	C	008
13	11R024	2	0.05	01	C	013
14	11R051	2	0.28	01	C	025
15	11R007	2	0.35	01	A	002
16	11R035	2	0.47	01	C	014

Grade 11 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	11R044	2	0.53	01	C	022
18	11R038	2	0.54	01	C	016
19	11R003	3	1.02	01	A	001
20	11R020	3	1.05	01	C	009
21	11R014	3	1.17	01	C	008
22	11R024	3	1.24	01	C	013
23	11R035	3	1.36	01	C	014
24	11R051	3	1.50	01	C	025
25	11R044	3	1.60	01	C	022
26	11R007	3	1.62	01	A	002
27	11R038	3	1.68	01	C	016
28	11R003	4	2.76	01	A	001
29	11R020	4	2.82	01	C	009
30	11R014	4	2.84	01	C	008
31	11R024	4	2.87	01	C	013
32	11R051	4	2.93	01	C	025
33	11R007	4	2.97	01	A	002

Grade 11 Reading Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
34	11R03 5	4	2.98	01	C	014
35	11R04 4	4	3.00	01	C	022
36	11R03 8	4	3.03	01	C	016

Grade 04 Science Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	4S01 4	1	-2.40	12	A	003
2	4S03 5	1	-2.33	12	E	040
3	4S04 2	1	-2.20	12	F	047
4	4S05 7	1	-1.99	13	B	013
5	4S00 8	1	-1.76	11	B	005
6	4S02 6	1	-1.73	12	C	017
7	4S03 5	2	-0.17	12	E	040
8	4S01 4	2	-0.12	12	A	003
9	4S04 2	2	-0.02	12	F	047
10	4S05 7	2	0.03	13	B	013
11	4S02 6	2	0.37	12	C	017
12	4S00 8	2	0.50	11	B	005
13	4S03 5	3	0.93	12	E	040
14	4S04 2	3	1.00	12	F	047
15	4S01 4	3	1.07	12	A	003
16	4S05 7	3	1.10	13	B	013

Grade 04 Science Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	4S026	3	1.37	12	C	017
18	4S008	3	1.52	11	B	005
19	4S042	4	2.42	12	F	047
20	4S035	4	2.44	12	E	040
21	4S014	4	2.45	12	A	003
22	4S057	4	2.45	13	B	013
23	4S008	4	2.65	11	B	005
24	4S026	4	2.69	12	C	017

Grade 07 Science Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	7S015	1	-2.76	12	A	001
2	7S003	1	-2.13	11	A	002
3	7S045	1	-1.92	12	F	091
4	7S052	1	-1.79	13	A	001
5	7S024	1	-1.72	12	C	049
6	7S037	1	-1.47	12	E	085
7	7S015	2	-0.49	12	A	001
8	7S003	2	-0.13	11	A	002
9	7S045	2	0.07	12	F	091
10	7S024	2	0.14	12	C	049
11	7S052	2	0.37	13	A	001
12	7S037	2	0.65	12	E	085
13	7S015	3	0.82	12	A	001
14	7S003	3	0.96	11	A	002
15	7S045	3	1.27	12	F	091
16	7S052	3	1.33	13	A	001

Grade 07 Science Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	7S02 4	3	1.36	12	C	049
18	7S03 7	3	1.63	12	E	085
19	7S01 5	4	2.40	12	A	001
20	7S00 3	4	2.51	11	A	002
21	7S04 5	4	2.55	12	F	091
22	7S02 4	4	2.59	12	C	049
23	7S05 2	4	2.69	13	A	001
24	7S03 7	4	2.78	12	E	085

Grade 11 Science Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	11S045	1	-2.78	13	A	001
2	11S023	1	-2.29	12	B	031
3	11S042	1	-1.76	12	F	102
4	11S008	1	-1.66	11	A	007
5	11S016	1	-1.50	12	A	025
6	11S035	1	-1.31	12	D	076
7	11S045	2	-0.43	13	A	001
8	11S023	2	-0.02	12	B	031
9	11S008	2	0.29	11	A	007
10	11S042	2	0.43	12	F	102
11	11S016	2	0.45	12	A	025
12	11S045	3	0.55	13	A	001
13	11S035	2	0.64	12	D	076
14	11S023	3	0.94	12	B	031
15	11S008	3	1.18	11	A	007
16	11S042	3	1.28	12	F	102

Grade 11 Science Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	11S016	3	1.44	12	A	025
18	11S035	3	1.50	12	D	076
19	11S045	4	2.08	13	A	001
20	11S023	4	2.23	12	B	031
21	11S008	4	2.37	11	A	007
22	11S042	4	2.41	12	F	102
23	11S016	4	2.52	12	A	025
24	11S035	4	2.53	12	D	076

Grade 06 Writing Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
1	6W02 1	1	-3.27	03	B,C	15
2	6W00 3	1	-2.24	03	A	1
3	6W01 8	1	-2.09	03	A	10
4	6W01 2	1	-1.94	03	A	9
5	6W03 0	1	-1.76	03	B,C	22
6	6W02 1	2	-0.87	03	B,C	15
7	6W01 8	2	0.02	03	A	10
8	6W00 3	2	0.05	03	A	1
9	6W01 2	2	0.17	03	A	9
10	6W02 1	3	0.45	03	B,C	15
11	6W03 0	2	0.48	03	B,C	22
12	6W00 3	3	1.12	03	A	1
13	6W01 8	3	1.20	03	A	10
14	6W01 2	3	1.21	03	A	9
15	6W03 0	3	1.55	03	B,C	22
16	6W02 1	4	1.76	03	B,C	15

Grade 06 Writing Item Map						
Page Number	Item	Score Point	Category Average	State Goal	State Standard	State Objective
17	6W003	4	2.30	03	A	1
18	6W018	4	2.31	03	A	10
19	6W012	4	2.44	03	A	9
20	6W030	4	2.71	03	B,C	22

Appendix F
2008 IAA Standard Setting Evaluation Form

**EVALUATION OF THE STANDARD SETTING WORKSHOP
FOR THE ILLINOIS ALTERNATIVE ASSESSMENTS**

GRADE 3-5 MATHEMATICS

May 14-16, 2008

The purpose of this evaluation is to secure your feedback about the standard setting process. Your feedback will provide a basis for evaluating the training, methods, and materials in the standard setting process.

Please complete the information below. Do not put your name on the form as we want your feedback to be anonymous.

1. Please read each of the following statements carefully. Place an X in one box for each statement to indicate the degree to which you agree with each statement.

	Strongly Agree	Agree	Disagree	Strongly disagree
a. I understood the purpose of this standard setting workshop.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. The feedback on cut scores gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. The feedback on impact gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. The description of the performance level descriptors was clear and understandable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. The description of the Body of Work Method was clear and understandable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. The description of the Item Mapping process was clear and understandable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. The description of the impact data was clear and understandable.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

2. Now that you have completed the standard setting process, how helpful were the following materials presented in the training in preparing you for the standard setting process.

	Strongly Agree	Agree	Disagree	Strongly disagree
a. The training materials contained all the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. The training on the content standards gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. The training on the Body of Work Method gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. The training on the Item Mapping process gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. The training on the Ordered Item Booklet gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. The training on the performance level descriptors gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. The training on the impact data gave me the information I needed to complete my assignment.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

3. Please rate the usefulness of the following materials or procedures in completing the standard setting process.

	Very useful	Somewhat useful	Not at all useful
a. Practicing the Body of Work Method	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Practicing the Item Mapping process	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Training materials	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Table discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Large group discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

4. How important was each of the following factors in placing your bookmark?

	Very important	Somewhat important	Not important
a. The description of performance level requirements	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Your perception of the difficulty of the items	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Your experiences with students	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Table discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Large group discussions	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Agreement feedback data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
g. Impact data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

5. Were any materials or procedures especially influential in your placement of the bookmark? If so, which ones? In what ways were they especially influential?

6. How appropriate was the amount of time you were given to complete the different components of the standard setting process?

	Too much	About right	Too little
a. Taking the assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. Scoring the assessment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. Training on the Body of Work method	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
d. Training on the Item Mapping process	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e. Table discussions on feedback	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
f. Group discussions on feedback	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

7. Please rate your level of comfort with the final cut scores.

	Very comfortable	Somewhat comfortable	Somewhat uncomfortable	Very uncomfortable
a. How comfortable are you with the final Grade 3 cut score?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
b. How comfortable are you with the final Grade 4 cut score?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
c. How comfortable are you with the final Grade 5 cut score?	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

8. After receiving the final cut scores and impact data, would you recommend changing the cut scores or leaving them unchanged (please explain any decisions to raise or lower a cut score)?

Grade 3:

☐ Leave the cut score as is
☐ Raise the cut score by points
☐ Lower the cut score by points

Grade 4:

☐ Leave the cut score as is
☐ Raise the cut score by points
☐ Lower the cut score by points

Grade 5:

☐ Leave the cut score as is
☐ Raise the cut score by points
☐ Lower the cut score by points

9. What suggestions do you have to improve the standard setting process and the training?
(Please use the reverse side as necessary.)

Appendix G
2008 IAA Standard Setting Evaluation Results

Table G.1. Standard Setting Evaluation Results – Lower Mathematics

Question	Mean	Median	SD
Scale: 1 = Strongly Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree			
I understood the purpose of this standard setting workshop.	1.27	1.00	0.47
The feedback on cut scores gave me the information I needed to complete my assignment.	1.36	1.00	0.50
The feedback on impact gave me the information I needed to complete my assignment.	1.36	1.00	0.50
The description of the performance level descriptors was clear and understandable.	1.27	1.00	0.47
The description of the Body of Work Method was clear and understandable.	1.45	1.00	0.52
The description of the Item Mapping process was clear and understandable.	1.45	1.00	0.52
The description of the impact data was clear and understandable.	1.18	1.00	0.40
The training materials contained all the information I needed to complete my assignment.	1.27	1.00	0.47
The training on the content standards gave me the information I needed to complete my assignment.	1.27	1.00	0.47
The training on the Body of Work Method gave me the information I needed to complete my assignment.	1.36	1.00	0.50
The training on the Item Mapping process gave me the information I needed to complete my assignment.	1.45	1.00	0.52
The training on the Ordered Item Booklet gave me the information I needed to complete my assignment.	1.45	1.00	0.52
The training on the performance level descriptors gave me the information I needed to complete my assignment.	1.27	1.00	0.47
The training on the impact data gave me the information I needed to complete my assignment.	1.36	1.00	0.50
Scale: 1 = Very Useful, 2 = Somewhat Useful, 3 = Not at all Useful			
Rate the usefulness of practicing the Body of Work Method	1.36	1.00	0.50
Rate the usefulness of practicing the Item Mapping process	1.27	1.00	0.65
Rate the usefulness of the training materials	1.27	1.00	0.47
Rate the usefulness of the table discussions	1.00	1.00	0.00
Rate the usefulness of the large group discussions	1.18	1.00	0.60

Question	Mean	Median	SD
Scale: 1 = Very Important, 2 = Somewhat Important, 3 = Not Important			
How important was the description of performance level requirements	1.18	1.00	0.40
How important was your perception of the difficulty of the items	1.18	1.00	0.40
How important was your experiences with students	1.00	1.00	0.00
How important were the table discussions	1.18	1.00	0.40
How important were the large group discussions	1.27	1.00	0.47
How important was the agreement feedback data	1.45	1.00	0.52
How important was the impact data	1.73	2.00	0.65
Scale: 1 = Too Much, 2 = About Right, 3 = Too Little			
How appropriate was the amount of time you were given taking the assessment	1.88	2.00	0.35
How appropriate was the amount of time you were given scoring the assessment	1.89	2.00	0.33
How appropriate was the amount of time you were given training on the Body of Work method	1.82	2.00	0.60
How appropriate was the amount of time you were given training on the Item Mapping process	2.00	2.00	0.63
How appropriate was the amount of time you were given for table discussions on feedback	2.09	2.00	0.30
How appropriate was the amount of time you were given for group discussions on feedback	2.00	2.00	0.00
Scale: 1 = Very Comfortable, 2 = Somewhat Comfortable, 3 = Somewhat Uncomfortable, 4 = Very Uncomfortable			
How comfortable are you with the final Grade 3 cut score?	1.90	1.50	1.10
How comfortable are you with the final Grade 4 cut score?	2.00	2.00	1.05
How comfortable are you with the final Grade 5 cut score?	1.80	1.00	1.14

Table G.2. Standard Setting Evaluation Results – Upper Mathematics

Question	Mean	Median	SD
Scale: 1 = Strongly Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree			
I understood the purpose of this standard setting workshop.	1.44	1.00	0.73
The feedback on cut scores gave me the information I needed to complete my assignment.	1.33	1.00	0.50
The feedback on impact gave me the information I needed to complete my assignment.	1.22	1.00	0.44
The description of the performance level descriptors was clear and understandable.	1.33	1.00	0.50
The description of the Body of Work Method was clear and understandable.	1.22	1.00	0.44
The description of the Item Mapping process was clear and understandable.	1.33	1.00	0.50
The description of the impact data was clear and understandable.	1.22	1.00	0.44
The training materials contained all the information I needed to complete my assignment.	1.22	1.00	0.44
The training on the content standards gave me the information I needed to complete my assignment.	1.33	1.00	0.50
The training on the Body of Work Method gave me the information I needed to complete my assignment.	1.22	1.00	0.44
The training on the Item Mapping process gave me the information I needed to complete my assignment.	1.33	1.00	0.50
The training on the Ordered Item Booklet gave me the information I needed to complete my assignment.	1.11	1.00	0.33
The training on the performance level descriptors gave me the information I needed to complete my assignment.	1.33	1.00	0.50
The training on the impact data gave me the information I needed to complete my assignment.	1.22	1.00	0.44
Scale: 1 = Very Useful, 2 = Somewhat Useful, 3 = Not at all Useful			
Rate the usefulness of practicing the Body of Work Method	1.33	1.00	0.71
Rate the usefulness of practicing the Item Mapping process	1.11	1.00	0.33
Rate the usefulness of the training materials	1.11	1.00	0.33
Rate the usefulness of the table discussions	1.00	1.00	0.00
Rate the usefulness of the large group discussions	1.00	1.00	0.00

Question	Mean	Median	SD
Scale: 1 = Very Important, 2 = Somewhat Important, 3 = Not Important			
How important was the description of performance level requirements	1.00	1.00	0.00
How important was your perception of the difficulty of the items	1.22	1.00	0.44
How important was your experiences with students	1.00	1.00	0.00
How important were the table discussions	1.11	1.00	0.33
How important were the large group discussions	1.00	1.00	0.00
How important was the agreement feedback data	1.11	1.00	0.33
How important was the impact data	1.11	1.00	0.33
Scale: 1 = Too Much, 2 = About Right, 3 = Too Little			
How appropriate was the amount of time you were given taking the assessment	2.00	2.00	0.00
How appropriate was the amount of time you were given scoring the assessment	2.00	2.00	0.00
How appropriate was the amount of time you were given training on the Body of Work method	1.89	2.00	0.33
How appropriate was the amount of time you were given training on the Item Mapping process	1.89	2.00	0.33
How appropriate was the amount of time you were given for table discussions on feedback	2.00	2.00	0.00
How appropriate was the amount of time you were given for group discussions on feedback	2.00	2.00	0.00
Scale: 1 = Very Comfortable, 2 = Somewhat Comfortable, 3 = Somewhat Uncomfortable, 4 = Very Uncomfortable			
How comfortable are you with the final Grade 6 cut score?	1.22	1.00	0.44
How comfortable are you with the final Grade 7 cut score?	1.22	1.00	0.44
How comfortable are you with the final Grade 8 cut score?	1.11	1.00	0.33
How comfortable are you with the final Grade 11 cut score?	1.11	1.00	0.33

Table G.3. Standard Setting Evaluation Results – Lower Reading

Question	Mean	Median	SD
Scale: 1 = Strongly Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree			
I understood the purpose of this standard setting workshop.	1.30	1.00	0.48
The feedback on cut scores gave me the information I needed to complete my assignment.	1.20	1.00	0.42
The feedback on impact gave me the information I needed to complete my assignment.	1.10	1.00	0.32
The description of the performance level descriptors was clear and understandable.	1.10	1.00	0.32
The description of the Body of Work Method was clear and understandable.	1.20	1.00	0.42
The description of the Item Mapping process was clear and understandable.	1.20	1.00	0.42
The description of the impact data was clear and understandable.	1.30	1.00	0.48
The training materials contained all the information I needed to complete my assignment.	1.20	1.00	0.42
The training on the content standards gave me the information I needed to complete my assignment.	1.00	1.00	0.00
The training on the Body of Work Method gave me the information I needed to complete my assignment.	1.00	1.00	0.00
The training on the Item Mapping process gave me the information I needed to complete my assignment.	1.10	1.00	0.32
The training on the Ordered Item Booklet gave me the information I needed to complete my assignment.	1.20	1.00	0.42
The training on the performance level descriptors gave me the information I needed to complete my assignment.	1.00	1.00	0.00
The training on the impact data gave me the information I needed to complete my assignment.	1.20	1.00	0.42
Scale: 1 = Very Useful, 2 = Somewhat Useful, 3 = Not at all Useful			
Rate the usefulness of practicing the Body of Work Method	1.20	1.00	0.42
Rate the usefulness of practicing the Item Mapping process	1.10	1.00	0.32
Rate the usefulness of the training materials	1.10	1.00	0.32
Rate the usefulness of the table discussions	1.10	1.00	0.32
Rate the usefulness of the large group discussions	1.00	1.00	0.00

Question	Mean	Median	SD
Scale: 1 = Very Important, 2 = Somewhat Important, 3 = Not Important			
How important was the description of performance level requirements	1.00	1.00	0.00
How important was your perception of the difficulty of the items	1.10	1.00	0.32
How important was your experiences with students	1.30	1.00	0.48
How important were the table discussions	1.10	1.00	0.32
How important were the large group discussions	1.00	1.00	0.00
How important was the agreement feedback data	1.20	1.00	0.42
How important was the impact data	1.20	1.00	0.42
Scale: 1 = Too Much, 2 = About Right, 3 = Too Little			
How appropriate was the amount of time you were given taking the assessment	1.90	2.00	0.32
How appropriate was the amount of time you were given scoring the assessment	1.90	2.00	0.32
How appropriate was the amount of time you were given training on the Body of Work method	2.00	2.00	0.00
How appropriate was the amount of time you were given training on the Item Mapping process	2.10	2.00	0.32
How appropriate was the amount of time you were given for table discussions on feedback	2.00	2.00	0.00
How appropriate was the amount of time you were given for group discussions on feedback	2.00	2.00	0.00
Scale: 1 = Very Comfortable, 2 = Somewhat Comfortable, 3 = Somewhat Uncomfortable, 4 = Very Uncomfortable			
How comfortable are you with the final Grade 3 cut score?	1.00	1.00	0.00
How comfortable are you with the final Grade 4 cut score?	1.13	1.00	0.35
How comfortable are you with the final Grade 5 cut score?	1.22	1.00	0.44

Table G.4. Standard Setting Evaluation Results – Upper Reading

Question	Mean	Median	SD
Scale: 1 = Strongly Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree			
I understood the purpose of this standard setting workshop.	1.10	1.00	0.32
The feedback on cut scores gave me the information I needed to complete my assignment.	1.10	1.00	0.32
The feedback on impact gave me the information I needed to complete my assignment.	1.10	1.00	0.32
The description of the performance level descriptors was clear and understandable.	1.10	1.00	0.32
The description of the Body of Work Method was clear and understandable.	1.40	1.00	0.70
The description of the Item Mapping process was clear and understandable.	1.00	1.00	0.00
The description of the impact data was clear and understandable.	1.30	1.00	0.48
The training materials contained all the information I needed to complete my assignment.	1.40	1.00	0.52
The training on the content standards gave me the information I needed to complete my assignment.	1.22	1.00	0.44
The training on the Body of Work Method gave me the information I needed to complete my assignment.	1.40	1.00	0.70
The training on the Item Mapping process gave me the information I needed to complete my assignment.	1.10	1.00	0.32
The training on the Ordered Item Booklet gave me the information I needed to complete my assignment.	1.10	1.00	0.32
The training on the performance level descriptors gave me the information I needed to complete my assignment.	1.10	1.00	0.32
The training on the impact data gave me the information I needed to complete my assignment.	1.10	1.00	0.32
Scale: 1 = Very Useful, 2 = Somewhat Useful, 3 = Not at all Useful			
Rate the usefulness of practicing the Body of Work Method	1.20	1.00	0.42
Rate the usefulness of practicing the Item Mapping process	1.10	1.00	0.32
Rate the usefulness of the training materials	1.10	1.00	0.32
Rate the usefulness of the table discussions	1.10	1.00	0.32
Rate the usefulness of the large group discussions	1.30	1.00	0.48

Question	Mean	Median	SD
Scale: 1 = Very Important, 2 = Somewhat Important, 3 = Not Important			
How important was the description of performance level requirements	1.00	1.00	0.00
How important was your perception of the difficulty of the items	1.10	1.00	0.32
How important was your experiences with students	1.25	1.00	0.71
How important were the table discussions	1.20	1.00	0.42
How important were the large group discussions	1.30	1.00	0.48
How important was the agreement feedback data	1.20	1.00	0.42
How important was the impact data	1.10	1.00	0.32
Scale: 1 = Too Much, 2 = About Right, 3 = Too Little			
How appropriate was the amount of time you were given taking the assessment	1.60	2.00	0.55
How appropriate was the amount of time you were given scoring the assessment	1.67	2.00	0.52
How appropriate was the amount of time you were given training on the Body of Work method	1.80	2.00	0.42
How appropriate was the amount of time you were given training on the Item Mapping process	1.80	2.00	0.42
How appropriate was the amount of time you were given for table discussions on feedback	1.80	2.00	0.42
How appropriate was the amount of time you were given for group discussions on feedback	1.80	2.00	0.42
Scale: 1 = Very Comfortable, 2 = Somewhat Comfortable, 3 = Somewhat Uncomfortable, 4 = Very Uncomfortable			
How comfortable are you with the final Grade 6 cut score?	1.20	1.00	0.42
How comfortable are you with the final Grade 7 cut score?	1.20	1.00	0.42
How comfortable are you with the final Grade 8 cut score?	1.20	1.00	0.42
How comfortable are you with the final Grade 11 cut score?	1.30	1.00	0.48

Table G.5. Standard Setting Evaluation Results - Science

Question	Mean	Median	SD
Scale: 1 = Strongly Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree			
I understood the purpose of this standard setting workshop.	1.11	1.00	0.33
The feedback on cut scores gave me the information I needed to complete my assignment.	1.00	1.00	0.00
The feedback on impact gave me the information I needed to complete my assignment.	1.11	1.00	0.33
The description of the performance level descriptors was clear and understandable.	1.56	1.00	1.01
The description of the Body of Work Method was clear and understandable.	1.44	1.00	0.73
The description of the Item Mapping process was clear and understandable.	1.44	1.00	0.53
The description of the impact data was clear and understandable.	1.00	1.00	0.00
The training materials contained all the information I needed to complete my assignment.	1.11	1.00	0.33
The training on the content standards gave me the information I needed to complete my assignment.	1.11	1.00	0.33
The training on the Body of Work Method gave me the information I needed to complete my assignment.	1.22	1.00	0.44
The training on the Item Mapping process gave me the information I needed to complete my assignment.	1.11	1.00	0.33
The training on the Ordered Item Booklet gave me the information I needed to complete my assignment.	1.33	1.00	0.50
The training on the performance level descriptors gave me the information I needed to complete my assignment.	1.44	1.00	1.01
The training on the impact data gave me the information I needed to complete my assignment.	1.13	1.00	0.35
Scale: 1 = Very Useful, 2 = Somewhat Useful, 3 = Not at all Useful			
Rate the usefulness of practicing the Body of Work Method	1.22	1.00	0.44
Rate the usefulness of practicing the Item Mapping process	1.00	1.00	0.00
Rate the usefulness of the training materials	1.22	1.00	0.44
Rate the usefulness of the table discussions	1.00	1.00	0.00
Rate the usefulness of the large group discussions	1.33	1.00	0.50

Question	Mean	Median	SD
Scale: 1 = Very Important, 2 = Somewhat Important, 3 = Not Important			
How important was the description of performance level requirements	1.22	1.00	0.44
How important was your perception of the difficulty of the items	1.22	1.00	0.44
How important was your experiences with students	1.22	1.00	0.44
How important were the table discussions	1.22	1.00	0.44
How important were the large group discussions	1.44	1.00	0.73
How important was the agreement feedback data	1.33	1.00	0.50
How important was the impact data	1.11	1.00	0.33
Scale: 1 = Too Much, 2 = About Right, 3 = Too Little			
How appropriate was the amount of time you were given taking the assessment	2.00	2.00	0.00
How appropriate was the amount of time you were given scoring the assessment	2.00	2.00	0.00
How appropriate was the amount of time you were given training on the Body of Work method	1.89	2.00	0.33
How appropriate was the amount of time you were given training on the Item Mapping process	2.00	2.00	0.00
How appropriate was the amount of time you were given for table discussions on feedback	2.00	2.00	0.00
How appropriate was the amount of time you were given for group discussions on feedback	1.89	2.00	0.33
Scale: 1 = Very Comfortable, 2 = Somewhat Comfortable, 3 = Somewhat Uncomfortable, 4 = Very Uncomfortable			
How comfortable are you with the final Grade 4 cut score?	1.00	1.00	0.00
How comfortable are you with the final Grade 7 cut score?	1.11	1.00	0.33
How comfortable are you with the final Grade 11 cut score?	1.33	1.00	0.71

Table G.6. Standard Setting Evaluation Results – Writing

Question	Mean	Median	SD
Scale: 1 = Strongly Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree			
I understood the purpose of this standard setting workshop.	1.00	1.00	0.00
The feedback on cut scores gave me the information I needed to complete my assignment.	1.00	1.00	0.00
The feedback on impact gave me the information I needed to complete my assignment.	1.00	1.00	0.00
The description of the performance level descriptors was clear and understandable.	1.00	1.00	0.00
The description of the Item Mapping process was clear and understandable.	1.00	1.00	0.00
The description of the impact data was clear and understandable.	1.00	1.00	0.00
The training materials contained all the information I needed to complete my assignment.	1.00	1.00	0.00
The training on the content standards gave me the information I needed to complete my assignment.	1.00	1.00	0.00
The training on the Item Mapping process gave me the information I needed to complete my assignment.	1.11	1.00	0.33
The training on the Ordered Item Booklet gave me the information I needed to complete my assignment.	1.11	1.00	0.33
The training on the performance level descriptors gave me the information I needed to complete my assignment.	1.11	1.00	0.33
The training on the impact data gave me the information I needed to complete my assignment.	1.11	1.00	0.33
Scale: 1 = Very Useful, 2 = Somewhat Useful, 3 = Not at all Useful			
Rate the usefulness of practicing the Item Mapping process	1.00	1.00	0.00
Rate the usefulness of the training materials	1.00	1.00	0.00
Rate the usefulness of the table discussions	1.11	1.00	0.33
Rate the usefulness of the large group discussions	1.00	1.00	0.00

Question	Mean	Median	SD
Scale: 1 = Very Important, 2 = Somewhat Important, 3 = Not Important			
How important was the description of performance level requirements	1.00	1.00	0.00
How important was your perception of the difficulty of the items	1.00	1.00	0.00
How important was your experiences with students	1.11	1.00	0.33
How important were the table discussions	1.33	1.00	0.50
How important were the large group discussions	1.22	1.00	0.44
How important was the impact data	1.00	1.00	0.00
Scale: 1 = Too Much, 2 = About Right, 3 = Too Little			
How appropriate was the amount of time you were given taking the assessment	1.71	2.00	0.49
How appropriate was the amount of time you were given scoring the assessment	1.71	2.00	0.49
How appropriate was the amount of time you were given training on the Item Mapping process	1.56	2.00	0.53
How appropriate was the amount of time you were given for table discussions on feedback	1.56	2.00	0.53
How appropriate was the amount of time you were given for group discussions on feedback	1.56	2.00	0.53
Scale: 1 = Very Comfortable, 2 = Somewhat Comfortable, 3 = Somewhat Uncomfortable, 4 = Very Uncomfortable			
How comfortable are you with the final Grade 6 cut score?	1.00	1.00	0.00

Appendix H

Item Analysis – Spring 2008 Field Test

Table H.1. Field Test Classical Item Statistics - Grade 3 Mathematics											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
03M003	210	10	6	A	1	0.77	3.27	13	11	12	64
03M009	149	9	6	A	5	0.64	2.93	12	19	32	37
03M020	155	9	6	B,C	9	0.73	3.06	16	15	14	54
03M028	159	10	6	B,C	10	0.68	2.65	21	28	16	35
03M035	149	10	7	A,B,C	5	0.70	3.24	11	13	16	60
03M040	155	10	8	A	1	0.79	3.40	9	10	12	68
03M053	196	9	9	A	1	0.77	2.98	19	15	14	52
03M057	182	10	9	B	9	0.67	3.08	17	13	16	54
03M062	199	10	10	A,B	1	0.76	3.28	12	11	15	63
3M017	196	10	6	A	7	0.70	2.95	16	18	21	45
3M032	182	9	7	A,B,C	1	0.62	2.35	25	33	24	18
3M033	199	9	7	A,B,C	5	0.62	3.58	7	6	9	78
3M045	159	9	8	C,D	4	0.70	2.57	25	23	22	30
3M063	210	9	10	A,B	1	0.62	2.70	20	23	24	33

Table H.2. Field Test Classical Item Statistics - Grade 4 Mathematics											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
04M003	166	10	6	A	1	0.45	2.66	19	31	13	36
04M008	176	10	6	A	2	0.67	3.03	13	20	19	48
04M011	171	9	6	A	5	0.79	3.20	12	13	18	57
04M017	253	10	6	A	8	0.78	3.18	12	14	17	57
04M023	246	10	6	B,C	11	0.65	2.87	17	23	15	44
04M039	172	10	7	A,B,C	4	0.65	3.00	12	22	20	46
04M048	176	9	8	C,D	7	0.65	2.77	14	30	22	34
04M050	171	10	9	A	1	0.68	3.20	11	15	18	56
04M058	166	9	9	B	12	0.69	3.22	13	13	13	61
04M062	197	10	10	A,B	1	0.73	3.15	10	18	18	53
04M064	253	9	10	A,B	1	0.73	3.33	10	11	15	64
4M020	172	9	6	B,C	10	0.71	2.73	16	28	24	33
4M034	246	9	7	A,B,C	2	0.72	3.01	16	16	17	50
4M041	197	9	8	A	1	0.71	3.27	9	14	18	59

Table H.3. Field Test Classical Item Statistics - Grade 5 Mathematics											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
05M004	178	10	6	A	1	0.57	3.00	10	22	25	43
05M008	252	9	6	A	7	0.67	3.11	10	20	21	50
05M014	161	10	6	B,C	12	0.73	2.89	20	19	14	48
05M033	202	10	7	A,B,C	4	0.64	3.56	30	10	14	73
05M044	252	10	8	C,D	8	0.70	3.01	13	19	22	46
05M045	186	9	8	C,D	8	0.75	2.70	24	16	27	33
05M054	190	10	9	B	14	0.57	3.47	40	12	17	67
05M058	186	10	10	A,B	1	0.80	3.09	13	17	18	52
5M009	171	10	6	A	10	0.75	3.04	18	13	18	51
5M018	190	9	6	B,C	13	0.67	2.98	13	25	13	49
5M031	161	9	7	A,B,C	2	0.71	3.08	16	16	12	56
5M036	171	9	8	A	1	0.76	2.96	19	18	11	52
5M050	202	9	9	A	1	0.62	3.30	81	11	22	58
5M064	178	9	10	C	5	0.59	3.01	10	20	29	41

Table H.4. Field Test Classical Item Statistics - Grade 6 Mathematics											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
06M001	237	9	6	A	1	0.66	3.28	8	16	16	60
06M024	268	10	7	A,B,C	3	0.67	3.24	11	13	17	59
06M036	204	9	8	B	6	0.72	3.08	14	14	23	50
06M042	237	10	8	C,D	9	0.67	3.13	11	18	19	52
06M046	202	10	9	A	2	0.74	3.22	12	11	21	56
06M052	204	10	9	B	12	0.80	3.49	8	9	9	74
06M059	202	9	10	A,B	1	0.74	3.12	10	19	21	50
06M060	182	10	10	A,B	1	0.68	2.86	16	21	22	40
06M065	268	9	10	C	6	0.70	3.30	9	13	18	61
06M101	217	10	6	B,C	12	0.73	3.18	9	18	17	55
6M007	192	10	6	A	4	0.66	2.86	17	23	17	43
6M017	217	9	7	A,B,C	1	0.68	3.30	8	12	22	58
6M040	182	9	8	C,D	9	0.67	2.85	14	26	21	39
6M050	192	9	9	B	9	0.74	3.26	13	14	10	64

Table H.5. Field Test Classical Item Statistics - Grade 7 Mathematics											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
07M001	192	10	6	A	3	0.67	3.30	7	17	16	60
07M010	293	9	6	B	8	0.70	2.94	15	21	20	44
07M016	205	10	6	D	16	0.78	3.25	10	16	12	61
07M021	180	9	7	A,B,C	1	0.71	3.44	7	10	16	67
07M027	211	9	7	A,B,C	3	0.61	3.20	9	12	29	50
07M028	196	10	8	A	1	0.69	2.87	12	29	21	39
07M041	180	10	8	C,D	12	0.58	2.66	13	36	24	27
07M048	205	9	9	B	11	0.72	3.21	14	11	16	60
07M053	213	10	9	B	14	0.71	3.35	12	10	9	69
07M058	293	10	10	A,B	1	0.74	3.26	11	16	11	63
07M060	196	9	10	A,B	1	0.75	2.99	14	18	24	44
07M063	213	9	10	C	7	0.75	3.04	14	17	20	49
7M037	192	9	8	B	8	0.64	2.92	14	23	20	43
7M044	211	10	9	A	5	0.63	2.80	15	24	28	33

Table H.6. Field Test Classical Item Statistics - Grade 8 Mathematics											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
08M007	279	10	6	A	3	0.68	3.22	11	15	15	59
08M011	194	9	6	B,C	9	0.72	3.24	9	14	20	57
08M019	200	10	7	A,B,C	1	0.74	3.32	14	10	8	69
08M022	210	10	7	A,B,C	1	0.75	3.26	10	16	12	62
08M024	201	9	7	A,B,C	3	0.65	2.76	13	28	29	30
08M031	226	10	8	A	1	0.76	3.23	11	15	14	60
08M041	201	10	8	C,D	13	0.75	3.13	12	15	21	52
08M048	210	9	9	B	10	0.77	3.31	8	13	19	60
08M051	194	10	9	B	10	0.59	3.19	11	16	15	57
08M054	220	10	10	A,B	1	0.82	3.18	11	19	11	59
08M059	279	9	10	A,B	3	0.71	3.30	7	16	17	60
8M013	226	9	6	D	17	0.58	2.68	18	24	30	28
8M034	220	9	8	B	8	0.75	3.14	12	15	20	53
8M063	200	9	10	C	7	0.73	3.47	8	9	13	71

Table H.7. Field Test Classical Item Statistics - Grade 11 Mathematics											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
11M004	260	10	6	A	1	0.75	3.37	9	10	15	66
11M006	150	9	6	B,C	9	0.75	3.06	11	15	31	43
11M011	148	9	6	B,C	10	0.73	3.27	8	13	23	56
11M014	170	10	6	B,C	13	0.81	3.34	14	9	8	70
11M015	133	10	6	B,C	13	0.52	3.56	6	6	14	74
11M032	106	9	7	A,B,C	3	0.64	2.76	21	21	20	39
11M038	181	9	8	A	4	0.65	3.21	10	12	25	53
11M039	106	10	8	A	4	0.73	3.15	13	17	11	58
11M047	181	10	9	A	2	0.67	3.13	10	14	27	49
11M050	170	9	9	A	6	0.73	3.45	6	11	15	68
11M055	150	10	9	A	7	0.76	3.37	11	7	18	65
11M063	260	9	10	A,B	1	0.64	2.84	12	25	30	33
11M064	148	10	10	A,B	1	0.79	3.26	10	14	17	59
11M101	133	9	8	B	12	0.48	3.23	11	9	26	54

Table H.8. Field Test Classical Item Statistics - Grade 3 Reading											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
03R003	211	8	1	A	3	0.73	2.87	17	18	27	39
03R010	182	8	1	A	8	0.71	2.78	19	21	21	38
03R012	196	8	1	A	11	0.73	3.04	16	17	14	53
03R020	159	8	1	B,C	14	0.72	3.36	9	13	13	66
03R028	155	9	1	B,C	18	0.69	2.89	19	15	25	41
03R034	151	8	1	C	23	0.53	3.15	7	21	21	50
03R041	199	9	2	A	7	0.79	3.35	8	11	19	62
03R042	151	9	2	A	7	0.62	2.91	12	23	27	38
03R047	159	9	2	B	10	0.65	2.58	23	26	21	30
03R050	196	9	2	B	10	0.65	2.67	25	19	19	37
03R100	211	9	1	B,C	13	0.77	3.06	16	13	20	51
03R101	182	9	1	B,C	20	0.57	2.47	23	30	26	21
3R026	199	8	1	B,C	18	0.74	3.36	9	10	18	63
3R038	155	8	1	C	27	0.74	3.14	10	14	26	49

Table H.9. Field Test Classical Item Statistics - Grade 4 Reading											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
04R002	253	8	1	A	4	0.64	3.14	11	16	23	51
04R009	166	8	1	B,C	7	0.75	3.27	12	10	17	61
04R014	197	8	1	B,C	9	0.80	3.45	7	8	18	67
04R019	246	8	1	B,C	10	0.63	2.88	13	25	21	40
04R023	166	9	1	B,C	15	0.65	2.91	16	22	16	46
04R024	197	9	1	C	17	0.72	3.24	10	15	16	59
04R026	246	9	1	C	17	0.71	2.93	19	17	16	48
04R033	171	9	1	C	21	0.69	2.95	14	18	27	41
04R038	171	8	1	C	25	0.59	2.94	13	20	25	42
04R043	172	8	2	A	1	0.68	3.04	9	24	19	47
04R046	176	8	2	A	6	0.66	2.83	18	22	20	40
04R048	253	9	2	A	6	0.65	3.07	13	16	21	49
04R051	176	9	2	B	13	0.66	2.72	19	25	20	36
04R053	172	9	2	B	13	0.61	2.58	19	33	22	27

Table H.10. Field Test Classical Item Statistics - Grade 5 Reading											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
05R001	252	8	1	A	2	0.73	3.10	10	20	19	51
05R004	189	9	1	A	2	0.48	3.14	8	20	21	51
05R008	189	8	1	B,C	7	0.60	2.98	11	25	20	44
05R015	161	8	1	B,C	8	0.60	2.91	16	25	11	48
05R021	161	9	1	B,C	13	0.67	2.91	15	19	27	39
05R027	171	8	1	C	16	0.69	2.92	15	21	20	44
05R029	202	9	1	C	20	0.62	3.03	13	18	21	48
05R033	178	9	1	C	20	0.56	3.10	10	18	26	47
05R036	186	8	1	C	26	0.78	2.82	18	24	16	42
05R047	202	8	2	A	1	0.67	3.53	4	12	10	74
05R052	178	8	2	A	5	0.62	3.17	10	15	24	52
05R055	186	9	2	A	8	0.74	2.65	20	26	21	32
05R064	171	9	2	B	14	0.68	2.73	18	24	25	33
5R018	252	9	1	B,C	13	0.73	2.94	15	18	24	43

Table H.11. Field Test Classical Item Statistics - Grade 6 Reading											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
06R001	268	8	1	A	3	0.67	3.06	10	19	24	47
06R009	192	9	1	A	5	0.58	2.78	14	28	25	33
06R019	237	9	1	B	11	0.66	3.39	5	12	21	62
06R023	217	9	1	B	12	0.71	3.28	9	13	18	59
06R027	201	8	1	B,C	13	0.77	3.05	12	17	25	46
06R032	268	9	1	C	16	0.65	2.89	12	26	22	40
06R034	192	8	1	C	16	0.70	3.01	11	23	20	46
06R038	204	8	1	C	20	0.70	3.01	14	16	25	45
06R045	182	8	1	C	22	0.70	3.22	12	13	17	58
06R049	181	9	2	A	7	0.71	3.13	10	17	23	50
06R057	217	8	2	A	9	0.65	3.16	8	22	15	54
06R061	237	8	2	A	14	0.61	2.96	11	21	30	38
06R100	204	9	2	A	1	0.72	3.14	14	10	23	53
6R010	201	9	1	A	5	0.74	3.39	10	9	13	68

Table H.12. Field Test Classical Item Statistics - Grade 7 Reading											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
07R001	294	8	1	A	3	0.73	3.01	14	19	21	47
07R009	192	9	1	A	5	0.66	3.30	7	16	18	59
07R015	205	9	1	B	9	0.73	3.37	10	11	12	67
07R016	294	9	1	B	9	0.76	2.96	15	19	20	46
07R019	180	8	1	B	14	0.69	3.14	9	18	22	51
07R026	211	8	1	C	15	0.62	3.20	9	15	22	54
07R030	180	9	1	C	17	0.66	2.97	10	22	28	39
07R033	192	8	1	C	19	0.56	3.02	7	22	32	39
07R040	205	8	1	C	22	0.63	3.23	9	12	25	54
07R044	197	8	2	A	1	0.68	3.14	9	18	24	49
07R050	213	8	2	A	6	0.61	3.02	13	19	23	46
07R055	211	9	2	A	8	0.67	2.94	16	17	26	42
07R058	213	9	2	A	8	0.62	2.84	15	23	24	38
7R028	197	9	1	C	17	0.67	3.04	12	19	22	47

Table H.13. Field Test Classical Item Statistics - Grade 8 Reading											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
08R004	278	8	1	A	3	0.62	2.93	12	22	26	40
08R009	201	9	1	B	8	0.82	3.41	9	8	14	68
08R013	210	9	1	B	8	0.75	3.09	11	18	23	49
08R018	221	9	1	B	10	0.80	3.21	12	14	14	59
08R021	201	8	1	B	12	0.76	2.92	15	20	22	42
08R026	210	8	1	C	14	0.73	3.33	9	13	13	65
08R027	278	9	1	C	14	0.80	3.50	6	9	12	73
08R030	201	9	1	C	16	0.70	2.91	17	19	18	45
08R035	221	8	1	C	18	0.81	3.47	8	10	10	72
08R043	201	8	1	C	23	0.71	3.19	13	11	19	57
08R049	194	9	2	A	4	0.59	2.98	8	26	24	41
08R061	194	8	2	A	13	0.53	3.11	9	21	20	51
08R100	226	8	2	A	1	0.74	2.99	13	20	20	46
8R052	226	9	2	A	6	0.69	2.94	13	24	19	44

Table H.14. Field Test Classical Item Statistics - Grade 11 Reading											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
11R002	262	8	1	A	1	0.81	3.47	9	6	14	71
11R011	170	8	1	A	2	0.75	3.31	12	9	13	65
11R021	170	9	1	C	9	0.58	3.05	14	14	26	46
11R023	134	8	1	C	9	0.53	3.34	7	11	22	60
11R025	134	9	1	C	13	0.58	3.72	4	4	8	84
11R026	181	8	1	C	13	0.70	3.34	10	10	15	64
11R030	181	9	1	C	14	0.68	3.32	9	12	17	62
11R033	106	8	1	C	14	0.74	3.11	11	14	26	48
11R041	148	8	1	C	16	0.78	3.39	8	9	20	64
11R045	148	9	1	C	22	0.78	3.56	6	8	9	76
11R047	150	9	1	C	22	0.81	3.55	7	6	13	74
11R050	150	8	1	C	25	0.82	3.49	6	9	14	71
11R054	106	9	1	C	16	0.75	3.25	11	9	23	57
11R100	262	9	1	C	8	0.67	3.11	11	11	33	45

Table H.15. Field Test Classical Item Statistics - Grade 4 Science											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
04S002	253	5	11	A	1	0.78	3.24	12	12	16	60
04S006	253	6	11	B	5	0.67	2.96	16	18	20	46
04S010	166	5	12	A	3	0.60	3.28	10	14	14	62
04S017	196	5	12	B	7	0.60	2.93	14	22	21	43
04S023	244	5	12	C	14	0.63	2.86	18	20	19	43
04S027	196	6	12	C	17	0.75	3.28	11	12	16	61
04S029	244	6	12	D	27	0.68	3.18	14	15	10	61
04S034	171	6	12	E	32	0.47	2.64	13	37	20	29
04S036	171	5	12	E	40	0.64	2.95	15	18	25	43
04S045	172	5	12	F	47	0.68	3.39	8	10	17	65
04S048	175	5	13	A	1	0.73	3.39	11	7	14	68
04S053	166	6	13	A	1	0.83	3.37	11	10	10	69
04S054	172	6	13	B	11	0.52	2.71	18	25	25	32
04S058	175	6	13	B	13	0.63	3.07	10	19	23	47

Table H.16. Field Test Classical Item Statistics - Grade 7 Science											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
07S007	292	5	11	A	2	0.82	3.28	11	13	12	64
07S016	180	5	12	A	1	0.71	3.58	6	9	6	79
07S020	211	6	12	A	15	0.65	3.26	9	11	26	54
07S023	191	6	12	C	49	0.79	3.29	9	14	16	61
07S028	205	6	12	C	53	0.58	3.20	8	17	20	54
07S031	196	6	12	D	69	0.61	2.82	18	23	17	41
07S032	213	6	12	E	76	0.63	2.95	18	17	17	48
07S035	180	6	12	E	85	0.61	2.81	12	29	24	34
07S039	211	5	12	E	88	0.54	3.18	9	16	25	51
07S041	191	5	12	E	88	0.58	2.98	9	22	31	38
07S047	205	5	12	F	91	0.66	3.01	13	21	17	49
07S049	196	5	12	F	91	0.60	2.65	18	28	23	30
07S054	292	6	13	A	1	0.75	3.23	11	12	19	58
07S056	213	5	13	B	11	0.72	3.14	13	13	22	52

Table H.17. Field Test Classical Item Statistics - Grade 11 Science											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
11S001	260	5	11	A	1	0.76	3.23	8	15	21	55
11S009	170	6	11	B	7	0.76	3.02	18	14	18	51
11S010	260	6	12	A	3	0.71	3.22	8	16	24	53
11S015	133	6	12	A	4	0.50	3.17	13	11	23	53
11S018	170	5	12	A	25	0.74	3.22	12	13	16	59
11S019	148	6	12	B	31	0.67	3.11	11	19	16	53
11S022	149	6	12	B	31	0.69	3.22	12	13	15	60
11S032	181	5	12	C	47	0.63	3.14	7	20	25	48
11S034	181	6	12	D	76	0.69	3.17	11	15	20	54
11S041	106	5	12	F	**	0.77	3.20	12	13	17	58
11S046	106	6	13	A	1	0.76	3.13	12	16	18	54
11S050	149	5	13	B	6	0.69	3.45	9	5	17	68
11S101	133	5	12	C	37	0.57	3.32	11	10	17	63
11S102	148	5	12	E	99	0.76	3.37	7	14	16	64

Table H.18. Field Test Classical Item Statistics - Grade 5 Writing											
								Percent at Score Point			
Item	N	Item Sequence	Task Goal	Task Standard	Task Objective	Item Total	Item Mean	1	2	3	4
5W019	186	4	3	B	15	0.80	2.93	19	15	22	45
5W035	202	4	3	B	21	0.77	3.27	11	11	17	60
5W101	161	4	3	A	9	0.77	2.91	17	22	13	48
5W115	170	4	3	A	10	0.76	2.96	18	18	15	49
5W131	190	4	3	A	1	0.65	3.42	6	11	19	64
5W136	177	4	3	B	21	0.70	3.28	9	12	21	58
5W202	252	4	3	A	1	0.68	3.46	7	12	11	71

Table H.19. Field Test Classical Item Statistics - Grade 6 Writing											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
6W102	268	4	3	A	1	0.66	2.72	13	31	25	31
6W103	202	4	3	A	9	0.68	2.98	15	17	22	46
6W104	192	4	3	A	10	0.64	2.84	17	22	21	40
6W106	204	4	3	B,C	15	0.81	3.32	13	9	12	67
6W108	217	4	3	B,C	22	0.61	3.29	10	14	13	63

Table H.19. Field Test Classical Item Statistics - Grade 6 Writing											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
6W109	236	4	3	B,C	37	0.67	3.39	8	11	15	66
6W119	182	4	3	B,C	15	0.66	3.12	9	18	24	48

Table H.20. Field Test Classical Item Statistics - Grade 8 Writing											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
8W059	200	4	3	B,C	15	0.70	2.95	16	21	17	47
8W101	278	4	3	A	1	0.76	3.45	8	9	12	71
8W102	200	4	3	A	1	0.82	3.50	8	8	11	74
8W103	210	4	3	A	9	0.76	3.16	15	9	21	55
8W107	221	4	3	A	10	0.79	3.17	10	20	12	57
8W111	194	4	3	B,C	22	0.56	3.07	8	24	23	46
8W113	226	4	3	B,C	38	0.76	3.17	12	16	15	57

Table H.21. Field Test Classical Item Statistics - Grade 11 Writing											
								Percent at Score Point			
Item	N	Item Sequence	State Goal	State Standard	State Objective	Item Total	Item Mean	1	2	3	4
9W103	170	4	3	A	12	0.74	3.11	15	14	16	55
9W202	261	4	3	A	9	0.84	3.48	9	7	11	73
9W205	134	4	3	A	21	0.57	3.61	5	5	13	77
9W208	181	4	3	B	28	0.73	3.39	9	11	12	68
9W211	106	4	3	B	32	0.78	3.29	13	6	20	61
9W213	150	4	3	B	34	0.77	3.47	5	9	19	67
9W216	148	4	3	B,C	35	0.79	3.55	7	5	14	74

Table H.22. DIF and Item Means by Gender and Ethnicity - Grade 3 Mathematics													
Item	Female		Male		White		Black		Hispanic		DIF		
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
03M003	67	3.15	143	3.33	111	3.33	37	2.92	40	3.30	0	0	0
03M009	59	3.10	90	2.82	66	3.02	36	2.97	37	2.76	0	0	0
03M020	54	2.87	101	3.17	84	3.07	28	3.18	26	2.92	0	0	0
03M028	50	2.60	109	2.67	82	2.74	33	2.64	35	2.57	0	0	0
03M035	59	3.36	90	3.17	66	3.21	36	3.44	37	3.16	0	0	0
03M040	54	3.19	101	3.51	84	3.45	28	3.29	26	3.15	0	0	0
03M053	75	2.93	121	3.02	107	3.22	37	2.78	39	2.56	0	0	0

Table H.22. DIF and Item Means by Gender and Ethnicity - Grade 3 Mathematics													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
03M057	63	3.06	119	3.08	96	3.18	48	2.90	25	2.84	0	0	0
03M062	68	3.25	131	3.30	119	3.39	40	2.75	24	3.50	0	0	0
3M017	75	2.83	121	3.02	107	3.15	37	2.84	39	2.56	0	0	0
3M032	63	2.40	119	2.32	96	2.44	48	2.10	25	2.36	0	0	0
3M033	68	3.65	131	3.55	119	3.61	40	3.35	24	3.67	0	0	0
3M045	50	2.26	109	2.72	82	2.49	33	2.76	35	2.63	0	0	0
3M063	67	2.72	143	2.70	111	2.68	37	2.81	40	2.58	0	0	0

Table H.23. DIF and Item Means by Gender and Ethnicity - Grade 4 Mathematics													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
04M003	56	2.66	110	2.66	101	2.66	26	2.04	23	3.04	0	0	0
04M008	66	3.03	110	3.03	84	3.20	56	2.89	31	2.81	0	0	0
04M011	58	2.98	113	3.31	90	3.22	49	3.12	23	3.22	0	0	0
04M017	78	2.94	175	3.29	146	3.12	39	3.41	48	3.17	1	0	0
04M023	90	2.72	156	2.95	146	2.90	44	2.75	41	2.73	0	0	0
04M039	57	3.00	115	3.00	98	3.02	44	3.05	22	2.82	0	0	0
04M048	66	2.59	110	2.87	84	2.90	56	2.68	31	2.68	0	0	0
04M050	58	2.97	113	3.32	90	3.24	49	3.31	23	2.74	0	0	1

Table H.23. DIF and Item Means by Gender and Ethnicity - Grade 4 Mathematics													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
04M058	56	3.23	110	3.21	101	3.30	26	2.65	23	3.43	0	0	0
04M062	74	3.22	123	3.11	111	3.18	38	3.21	37	3.03	0	0	0
04M064	78	3.28	175	3.35	146	3.32	39	3.38	48	3.31	0	0	0
4M020	57	2.61	115	2.79	98	2.68	44	2.98	22	2.45	0	0	0
4M034	90	2.81	156	3.13	146	3.01	44	2.93	41	3.02	0	0	0
4M041	74	3.35	123	3.22	111	3.21	38	3.53	37	3.24	0	0	0

Table H.24. DIF and Item Means by Gender and Ethnicity - Grade 5 Mathematics													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
05M004	68	2.91	110	3.05	99	2.97	44	3.14	22	3.00	0	1	0
05M008	90	3.14	162	3.09	130	3.19	57	3.04	49	3.02	0	0	0
05M014	55	3.02	106	2.83	105	2.97	21	2.33	25	2.92	0	0	0
05M033	72	3.69	130	3.48	117	3.51	36	3.50	39	3.67	0	0	0
05M044	90	2.97	162	3.03	130	2.95	57	3.07	49	3.08	0	0	0
05M045	65	2.69	121	2.70	86	2.84	41	2.66	40	2.65	0	0	0
05M054	70	3.43	120	3.50	117	3.55	45	3.42	17	3.06	0	0	0
05M058	65	3.02	121	3.13	86	3.21	41	3.10	40	2.88	0	0	0
5M009	66	2.98	105	3.07	79	3.14	44	2.70	31	3.16	0	0	0
5M018	70	3.04	120	2.95	117	3.09	45	2.87	17	2.47	1	0	0
5M031	55	3.16	106	3.04	105	3.04	21	2.86	25	3.40	0	0	0
5M036	66	3.12	105	2.87	79	3.19	44	2.55	31	3.00	0	0	0
5M050	72	3.38	130	3.25	117	3.27	36	3.11	39	3.49	0	0	0
5M064	68	2.81	110	3.13	99	3.09	44	2.91	22	3.05	0	0	0

Table H.25. DIF and Item Means by Gender and Ethnicity - Grade 6 Mathematics													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
06M001	87	3.37	150	3.23	137	3.40	50	3.02	31	3.10	0	0	0
06M024	103	3.10	165	3.33	149	3.26	65	3.18	41	3.29	0	0	0
06M036	65	3.05	139	3.09	113	3.04	51	3.14	31	3.29	0	0	0
06M042	87	3.28	150	3.05	137	3.26	50	2.84	31	3.16	0	0	0
06M046	70	3.04	132	3.31	97	3.41	50	3.34	45	2.69	0	0	0
06M052	65	3.34	139	3.55	113	3.48	51	3.59	31	3.48	0	0	0
06M059	70	3.03	132	3.17	97	3.31	50	3.16	45	2.73	0	0	0
06M060	62	2.90	120	2.83	105	3.05	47	2.62	22	2.59	0	0	0
06M065	103	3.16	165	3.39	149	3.34	65	3.42	41	3.10	0	0	1
06M101	73	3.14	144	3.21	116	3.15	52	3.12	40	3.38	0	0	0
6M007	73	2.92	119	2.82	101	2.91	39	2.49	42	3.05	0	0	0
6M017	73	3.27	144	3.31	116	3.41	52	3.08	40	3.35	0	0	0
6M040	62	2.84	120	2.85	105	2.94	47	2.68	22	2.82	0	0	0
6M050	73	3.19	119	3.29	101	3.29	39	2.79	42	3.62	0	0	0

Table H.26. DIF and Item Means by Gender and Ethnicity - Grade 7 Mathematics													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
07M001	83	3.36	109	3.25	98	3.39	48	3.00	32	3.25	0	0	0
07M010	90	2.92	203	2.95	163	3.12	64	2.78	48	2.56	0	0	0
07M016	85	3.29	120	3.22	118	3.20	51	3.25	29	3.45	0	0	0
07M021	66	3.56	114	3.37	89	3.53	23	3.30	57	3.33	0	0	0
07M027	82	3.24	129	3.18	98	3.18	42	3.10	53	3.30	0	0	0
07M028	71	2.87	125	2.86	97	2.86	59	3.02	30	2.97	0	0	0
07M041	66	2.65	114	2.67	89	2.63	23	2.39	57	2.82	0	0	0
07M048	85	3.13	120	3.27	118	3.19	51	3.16	29	3.41	0	0	0
07M053	81	3.46	132	3.28	115	3.52	47	3.04	38	3.11	0	0	0
07M058	90	3.19	203	3.30	163	3.34	64	3.00	48	3.33	0	0	0
07M060	71	3.07	125	2.94	97	3.03	59	3.05	30	3.03	0	0	0
07M063	81	3.15	132	2.98	115	3.17	47	2.77	38	3.05	0	0	0
7M037	83	3.06	109	2.82	98	3.00	48	2.77	32	2.97	0	0	0
7M044	82	2.76	129	2.82	98	2.95	42	2.45	53	2.72	0	0	0

Table H.27. DIF and Item Means by Gender and Ethnicity - Grade 8 Mathematics													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
08M007	114	3.22	165	3.22	157	3.25	74	3.12	38	3.37	0	0	0
08M011	81	3.19	113	3.27	101	3.33	50	3.26	34	2.85	0	0	0
08M019	85	3.38	115	3.27	109	3.48	52	2.81	31	3.42	0	0	0
08M022	88	3.26	122	3.25	95	3.21	64	3.28	41	3.29	0	0	0
08M024	67	2.63	134	2.83	103	2.86	54	2.61	30	2.73	0	0	0
08M031	91	3.36	135	3.15	125	3.26	52	3.27	35	3.23	0	0	1
08M041	67	3.00	134	3.20	103	3.24	54	2.93	30	3.13	0	0	0
08M048	88	3.35	122	3.29	95	3.24	64	3.23	41	3.61	0	0	1
08M051	81	3.11	113	3.24	101	3.24	50	3.06	34	3.12	0	0	0
08M054	74	3.14	146	3.21	123	3.20	56	3.11	35	3.20	0	0	0
08M059	114	3.28	165	3.31	157	3.30	74	3.27	38	3.34	0	0	0
8M013	91	2.58	135	2.75	125	2.59	52	2.90	35	2.83	0	0	0
8M034	74	3.27	146	3.07	123	3.21	56	3.05	35	3.00	0	0	0
8M063	85	3.61	115	3.36	109	3.60	52	3.06	31	3.58	0	0	0

Table H.28. DIF and Item Means by Gender and Ethnicity - Grade 11 Mathematics													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
11M004	99	3.31	161	3.41	175	3.36	46	3.37	26	3.23	0	0	0
11M006	63	3.17	87	2.98	61	3.20	74	3.04	13	2.77	0	0	0
11M011	47	3.13	101	3.34	93	3.37	31	3.06	18	3.28	0	1	0
11M014	63	3.48	107	3.26	100	3.57	36	2.86	23	3.35	0	0	0
11M015	54	3.46	79	3.62	89	3.57	27	3.67	12	3.08	0	0	1
11M032	39	2.85	67	2.72	61	2.93	28	2.61	14	2.43	0	0	0
11M038	80	3.18	101	3.24	88	3.24	45	3.04	43	3.42	0	0	0
11M039	39	3.31	67	3.06	61	3.31	28	2.89	14	3.07	0	0	0
11M047	80	3.04	101	3.21	88	3.13	45	2.96	43	3.44	0	0	0
11M050	63	3.63	107	3.34	100	3.62	36	2.94	23	3.57	0	0	0
11M055	63	3.49	87	3.28	61	3.44	74	3.42	13	3.00	0	0	0
11M063	99	2.84	161	2.84	175	2.89	46	2.80	26	2.65	0	0	0
11M064	47	3.21	101	3.28	93	3.33	31	3.26	18	2.94	0	0	0
11M101	54	3.11	79	3.30	89	3.25	27	3.11	12	3.25	0	0	0

Table H.29. DIF and Item Means by Gender and Ethnicity - Grade 3 Reading													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
03R003	67	2.84	144	2.89	112	2.88	37	2.86	40	2.78	0	0	0
03R010	63	2.97	119	2.68	96	2.94	48	2.58	25	2.56	1	0	0
03R012	75	3.09	121	3.01	107	3.32	37	2.78	39	2.62	1	0	0
03R020	50	3.34	109	3.37	82	3.44	33	3.64	35	3.11	0	0	0
03R028	54	2.85	101	2.91	84	2.94	28	2.89	26	2.50	0	0	0
03R034	60	3.25	91	3.08	67	3.09	36	3.44	38	3.05	0	0	0
03R041	68	3.37	131	3.34	119	3.43	40	2.93	24	3.46	0	0	1
03R042	60	3.07	91	2.80	67	3.04	36	3.03	38	2.74	0	0	0
03R047	50	2.48	109	2.62	82	2.73	33	2.36	35	2.43	0	0	0
03R050	75	2.64	121	2.69	107	2.75	37	2.51	39	2.69	0	0	1
03R100	67	2.99	144	3.10	112	3.14	37	2.81	40	2.98	0	0	0
03R101	63	2.68	119	2.35	96	2.59	48	2.25	25	2.56	1	0	0
3R026	68	3.37	131	3.35	119	3.49	40	2.85	24	3.33	0	0	0
3R038	54	3.17	101	3.13	84	3.10	28	3.11	26	3.04	0	0	0

Table H.30. DIF and Item Means by Gender and Ethnicity - Grade 4 Reading													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
04R002	78	3.14	175	3.14	146	3.15	39	3.15	48	3.00	0	0	0
04R009	56	3.34	110	3.23	101	3.40	26	2.69	23	3.30	0	0	1
04R014	74	3.62	123	3.34	111	3.45	38	3.58	37	3.38	1	0	0
04R019	90	2.83	156	2.91	146	2.89	44	2.50	41	3.15	0	0	0
04R023	56	3.02	110	2.85	101	2.84	26	2.58	23	3.39	0	0	0
04R024	74	3.32	123	3.20	111	3.14	38	3.47	37	3.30	0	0	0
04R026	90	2.77	156	3.02	146	3.05	44	2.39	41	3.00	0	1	0
04R033	58	2.79	113	3.04	90	3.08	49	2.88	23	2.61	0	0	1
04R038	58	2.81	113	3.01	90	3.03	49	2.84	23	2.78	0	0	0
04R043	57	3.16	115	2.98	98	3.06	44	3.11	22	2.77	0	0	0
04R046	66	2.71	110	2.90	84	2.90	56	2.79	31	2.84	0	0	0
04R048	78	2.94	175	3.13	146	3.15	39	3.10	48	2.88	0	0	1
04R051	66	2.70	110	2.74	84	2.76	56	2.70	31	2.74	0	0	0
04R053	57	2.47	115	2.63	98	2.53	44	2.82	22	2.18	0	1	0

Table H.31. DIF and Item Means by Gender and Ethnicity - Grade 5 Reading													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
05R001	90	3.00	162	3.16	130	3.09	57	3.14	49	3.12	1	0	0
05R004	70	3.16	119	3.13	116	3.12	45	3.22	17	3.00	0	0	0
05R008	70	2.87	119	3.04	116	2.97	45	3.11	17	2.53	0	0	0
05R015	55	2.87	106	2.93	105	3.03	21	2.33	25	2.84	0	0	0
05R021	55	2.98	106	2.87	105	2.97	21	2.52	25	3.12	0	0	0
05R027	66	2.91	105	2.93	79	3.05	44	2.66	31	2.77	0	0	0
05R029	72	3.22	130	2.93	117	3.02	36	3.00	39	3.23	0	0	0
05R033	68	3.01	110	3.15	99	3.20	44	2.93	22	3.27	0	0	0
05R036	65	2.78	121	2.84	86	2.99	41	2.76	40	2.60	0	0	0
05R047	72	3.65	130	3.46	117	3.48	36	3.56	39	3.62	0	0	0
05R052	68	2.90	110	3.35	99	3.25	44	3.11	22	3.27	1	0	0
05R055	65	2.58	121	2.69	86	2.90	41	2.68	40	2.28	0	0	0
05R064	66	2.74	105	2.72	79	2.85	44	2.52	31	2.58	0	0	0
5R018	90	2.99	162	2.92	130	2.94	57	2.72	49	3.29	0	0	0

Table H.32. DIF and Item Means by Gender and Ethnicity - Grade 6 Reading													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
06R001	103	2.82	165	3.22	149	3.01	65	3.22	41	3.00	0	0	0
06R009	73	2.74	119	2.81	101	2.80	39	2.54	42	2.90	0	0	0
06R019	87	3.52	150	3.31	137	3.49	50	3.20	31	3.23	0	0	0
06R023	73	3.12	144	3.35	116	3.35	52	3.15	40	3.23	0	0	0
06R027	70	2.93	131	3.12	97	3.28	50	3.12	44	2.48	0	0	0
06R032	103	2.73	165	2.99	149	2.82	65	2.92	41	3.12	0	0	0
06R034	73	2.99	119	3.02	101	3.15	39	2.69	42	2.90	0	0	0
06R038	65	3.12	139	2.96	113	3.05	51	2.96	31	3.10	0	0	0
06R045	62	3.18	120	3.24	105	3.37	47	3.06	22	2.86	0	0	0
06R049	62	3.15	119	3.12	105	3.32	46	2.87	22	2.82	0	0	0
06R057	73	3.11	144	3.18	116	3.28	52	2.90	40	3.15	0	0	0
06R061	87	3.00	150	2.93	137	3.05	50	2.82	31	2.81	0	0	0
06R100	65	3.09	139	3.17	113	3.21	51	3.18	31	3.16	0	0	0
6R010	70	3.33	131	3.43	97	3.65	50	3.50	44	2.84	0	0	0

Table H.33. DIF and Item Means by Gender and Ethnicity - Grade 7 Reading													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
07R001	91	3.02	203	3.00	164	3.20	64	2.64	48	2.75	0	0	0
07R009	83	3.20	109	3.37	98	3.44	48	2.94	32	3.25	1	0	0
07R015	85	3.33	120	3.39	118	3.38	51	3.39	29	3.21	0	0	0
07R016	91	3.05	203	2.91	164	3.09	64	2.67	48	2.83	0	0	0
07R019	66	3.26	114	3.08	89	3.13	23	3.04	57	3.16	0	0	0
07R026	82	3.20	129	3.21	98	3.20	42	3.12	53	3.23	0	0	0
07R030	66	3.00	114	2.96	89	2.98	23	2.78	57	2.98	0	0	0
07R033	83	3.04	109	3.00	98	3.06	48	2.90	32	3.00	0	0	0
07R040	85	3.25	120	3.23	118	3.21	51	3.20	29	3.38	0	0	0
07R044	71	3.18	126	3.11	97	3.20	59	3.20	30	3.23	0	0	0
07R050	81	3.12	132	2.95	115	3.16	47	2.79	38	2.97	0	0	0
07R055	82	2.90	129	2.96	98	3.03	42	2.81	53	2.87	0	0	0
07R058	81	2.90	132	2.80	115	2.97	47	2.47	38	2.84	0	0	0
7R028	71	3.08	126	3.02	97	3.12	59	3.07	30	2.90	0	0	1

Table H.34. DIF and Item Means by Gender and Ethnicity - Grade 8 Reading													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
08R004	113	2.95	165	2.92	156	3.06	74	2.78	38	2.76	0	0	1
08R009	85	3.41	116	3.41	109	3.59	52	2.98	32	3.38	0	0	0
08R013	88	3.06	122	3.11	95	3.02	64	3.11	41	3.27	0	0	0
08R018	74	3.16	147	3.23	123	3.30	56	3.05	36	3.11	1	0	0
08R021	85	2.96	116	2.89	109	3.16	52	2.35	32	2.88	0	1	0
08R026	88	3.19	122	3.43	95	3.28	64	3.33	41	3.46	0	0	0
08R027	113	3.46	165	3.53	156	3.54	74	3.38	38	3.58	0	0	0
08R030	67	2.82	134	2.96	103	3.17	54	2.46	30	2.77	0	1	0
08R035	74	3.50	147	3.46	123	3.68	56	3.20	36	3.17	0	1	2
08R043	67	3.25	134	3.16	103	3.32	54	3.00	30	3.10	0	0	0
08R049	81	2.98	113	2.99	101	3.00	50	3.16	34	2.82	0	0	0
08R061	81	3.07	113	3.14	101	3.11	50	3.30	34	2.85	0	0	0
08R100	91	3.09	135	2.93	125	2.94	52	3.19	35	3.03	0	0	0
8R052	91	3.10	135	2.84	125	2.88	52	3.19	35	2.94	0	0	0

Table H.35. DIF and Item Means by Gender and Ethnicity - Grade 11 Reading													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
11R00 2	100	3.38	16 2	3.52	17 7	3.46	46	3.37	26	3.38	0	0	0
11R01 1	63	3.63	10 7	3.12	10 0	3.49	36	2.89	23	3.22	1	0	0
11R02 1	63	3.02	10 7	3.07	10 0	3.29	36	2.61	23	2.52	0	0	1
11R02 3	54	3.30	80	3.36	90	3.38	27	3.19	12	3.50	0	0	0
11R02 5	54	3.59	80	3.80	90	3.70	27	3.67	12	3.83	0	0	0
11R02 6	80	3.26	10 1	3.40	88	3.42	45	3.27	43	3.35	0	0	0
11R03 0	80	3.28	10 1	3.36	88	3.32	45	3.07	43	3.58	0	0	0
11R03 3	39	3.03	67	3.16	61	3.38	28	2.64	14	3.07	0	1	0
11R04 1	47	3.26	10 1	3.45	93	3.40	31	3.42	18	3.39	0	0	0
11R04 5	47	3.34	10 1	3.66	93	3.67	31	3.32	18	3.56	0	1	0
11R04 7	63	3.57	87	3.53	61	3.59	74	3.64	13	3.15	0	0	0
11R05 0	63	3.63	87	3.39	61	3.52	74	3.57	13	3.15	0	0	0
11R05 4	39	3.28	67	3.22	61	3.38	28	3.11	14	3.00	0	1	0
11R10 0	100	3.11	16 2	3.10	17 7	3.10	46	3.17	26	2.88	0	0	0

Table H.36. DIF and Item Means by Gender and Ethnicity - Grade 4 Science													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
04S002	78	3.31	175	3.21	146	3.15	39	3.46	48	3.31	0	0	0
04S006	78	2.92	175	2.98	146	2.99	39	3.10	48	2.67	0	0	2
04S010	56	3.45	110	3.19	101	3.23	26	3.08	23	3.48	0	0	0
04S017	74	2.92	122	2.94	111	2.87	37	3.19	37	2.81	0	0	0
04S023	89	2.70	155	2.96	145	2.93	43	2.60	41	2.95	0	0	0
04S027	74	3.32	122	3.25	111	3.32	37	3.49	37	3.00	0	0	0
04S029	89	2.99	155	3.29	145	3.32	43	2.86	41	3.12	0	0	0
04S034	58	2.71	113	2.61	90	2.73	49	2.45	23	2.35	1	0	0
04S036	58	2.81	113	3.03	90	2.98	49	2.82	23	3.04	0	0	0
04S045	57	3.35	115	3.41	98	3.42	44	3.39	22	3.36	0	0	0
04S048	65	3.37	110	3.41	84	3.50	56	3.21	30	3.53	0	0	0
04S053	56	3.55	110	3.27	101	3.45	26	2.81	23	3.61	0	0	0
04S054	57	2.68	115	2.72	98	2.68	44	2.75	22	2.68	0	0	0
04S058	65	3.06	110	3.08	84	3.23	56	2.84	30	3.27	0	0	0

Table H.37. DIF and Item Means by Gender and Ethnicity - Grade 7 Science													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
07S007	90	3.47	202	3.20	162	3.43	64	3.00	48	3.19	0	0	0
07S016	66	3.68	114	3.53	89	3.61	23	3.52	57	3.51	0	0	0
07S020	82	3.18	129	3.30	98	3.19	42	3.31	53	3.25	0	0	0
07S023	82	3.40	109	3.21	98	3.48	47	2.98	32	3.00	0	0	0
07S028	85	3.20	120	3.21	118	3.27	51	3.08	29	3.17	0	0	0
07S031	71	2.80	125	2.82	97	2.89	59	2.85	30	2.87	0	0	0
07S032	81	3.04	132	2.89	115	3.14	47	2.68	38	2.68	0	0	0
07S035	66	2.73	114	2.86	89	2.79	23	2.78	57	2.82	0	0	0
07S039	82	3.10	129	3.23	98	3.28	42	2.93	53	3.23	0	0	0
07S041	82	3.13	109	2.86	98	3.10	47	2.94	32	2.88	0	0	0
07S047	85	3.00	120	3.02	118	2.99	51	3.24	29	2.62	0	0	0
07S049	71	2.77	125	2.58	97	2.76	59	2.56	30	2.67	0	0	0
07S054	90	3.33	202	3.18	162	3.31	64	2.98	48	3.15	0	0	0
07S056	81	3.10	132	3.16	115	3.28	47	2.89	38	3.03	0	0	0

Table H.38. DIF and Item Means by Gender and Ethnicity - Grade 11 Science													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
11S001	99	3.21	161	3.24	175	3.19	46	3.33	26	3.08	0	0	0
11S009	63	3.14	107	2.94	100	3.30	36	2.42	23	2.83	0	1	2
11S010	99	3.26	161	3.19	175	3.20	46	3.35	26	2.96	1	0	0
11S015	54	3.09	79	3.23	89	3.19	27	3.33	12	2.92	0	0	0
11S018	63	3.29	107	3.19	100	3.38	36	2.94	23	3.17	0	0	0
11S019	47	2.87	101	3.23	93	3.20	31	2.87	18	3.17	0	0	0
11S022	62	3.19	87	3.24	61	3.23	73	3.32	13	2.92	0	0	0
11S032	80	3.08	101	3.19	88	3.27	45	3.11	43	2.88	0	0	0
11S034	80	3.16	101	3.18	88	3.48	45	2.93	43	2.84	0	1	1
11S041	39	3.23	67	3.18	61	3.41	28	2.96	14	2.93	0	0	0
11S046	39	3.18	67	3.10	61	3.26	28	2.96	14	3.07	0	0	0
11S050	62	3.50	87	3.41	61	3.41	73	3.53	13	3.38	0	0	0
11S101	54	3.04	79	3.52	89	3.36	27	3.30	12	2.83	0	0	0
11S102	47	3.21	101	3.45	93	3.47	31	3.26	18	3.11	0	0	0

Table H.39. DIF and Item Means by Gender and Ethnicity - Grade 5 Writing													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
5W019	65	2.89	121	2.95	86	3.07	41	2.85	40	2.85	0	0	0
5W035	72	3.29	130	3.26	117	3.28	36	3.25	39	3.18	1	0	0
5W101	55	3.02	106	2.86	105	3.00	21	2.67	25	2.72	0	0	1
5W115	66	2.85	104	3.03	79	3.10	43	2.67	31	3.03	0	0	0
5W131	70	3.36	120	3.45	117	3.41	45	3.53	17	3.12	0	0	0
5W136	68	3.16	109	3.35	99	3.35	44	3.16	21	3.43	0	0	0
5W202	90	3.51	162	3.43	130	3.44	57	3.37	49	3.71	0	0	0

Table H.40. DIF and Item Means by Gender and Ethnicity - Grade 6 Writing													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
6W10 2	103	2.59	16 5	2.81	14 9	2.73	65	2.83	41	2.71	0	0	1
6W10 3	70	2.99	13 2	2.98	97	3.29	50	3.12	45	2.33	0	0	0
6W10 4	73	2.90	11 9	2.81	10 1	2.93	39	2.59	42	2.76	0	0	0
6W10 6	65	3.23	13 9	3.37	11 3	3.32	51	3.53	31	3.23	0	0	0
6W10 8	73	3.38	14 4	3.25	11 6	3.44	52	3.12	40	3.13	0	0	0
6W10 9	86	3.47	15 0	3.34	13 6	3.41	50	3.36	31	3.35	0	0	0
6W11 9	62	3.10	12 0	3.13	10 5	3.36	47	2.79	22	2.68	0	1	0

Table H.41. DIF and Item Means by Gender and Ethnicity - Grade 8 Writing													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
8W05 9	67	2.85	13 3	2.99	10 2	3.08	54	2.65	30	2.90	0	0	0
8W10 1	113	3.54	16 5	3.38	15 6	3.49	74	3.36	38	3.39	0	0	0
8W10 2	85	3.61	11 5	3.41	10 9	3.69	52	3.04	31	3.45	0	0	0
8W10 3	88	3.20	12 2	3.13	95	3.18	64	3.06	41	3.20	0	0	0
8W10 7	74	3.30	14 7	3.10	12 3	3.31	56	2.98	36	2.97	0	0	0
8W11 1	81	2.99	11 3	3.12	10 1	3.13	50	3.14	34	2.74	0	0	0
8W11 3	91	3.29	13 5	3.10	12 5	3.22	52	3.15	35	3.20	0	0	0

Table H.42. DIF and Item Means by Gender and Ethnicity - Grade 11 Writing													
	Female		Male		White		Black		Hispanic		DIF		
Item	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	Female	Black	Hispanic
9W103	63	3.16	107	3.07	100	3.28	36	2.69	23	3.13	0	0	0
9W202	100	3.49	161	3.48	176	3.49	46	3.46	26	3.31	0	0	0
9W205	54	3.46	80	3.71	90	3.66	27	3.67	12	3.00	0	0	2
9W208	80	3.41	101	3.38	88	3.58	45	3.20	43	3.35	0	0	0
9W211	39	3.33	67	3.27	61	3.44	28	3.11	14	3.14	0	0	0
9W213	63	3.59	87	3.39	61	3.54	74	3.50	13	3.23	1	1	0
9W216	47	3.47	101	3.59	93	3.59	31	3.52	18	3.50	0	0	0

Table H.43. IRT Item Analysis Results – Grade 3 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
03M003	210	0.93	-0.99	0.28	-0.45	-2.71	-0.13	0.66	1.71
03M009	149	0.92	-0.40	0.45	1.54	-0.93	0.57	1.39	2.16
03M020	155	0.94	-0.17	0.89	0.25	-1.43	0.61	1.00	2.13
03M028	159	0.95	-0.44	1.49	0.83	-1.62	0.62	1.33	2.12
03M035	149	0.82	-0.20	0.49	0.11	-1.08	0.44	0.97	1.91
03M040	155	0.80	-1.10	0.24	-0.41	-2.39	-0.14	0.71	1.91
03M053	196	0.84	-0.35	0.63	0.17	-1.65	0.02	1.00	2.04

Table H.43. IRT Item Analysis Results – Grade 3 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
03M05 7	18 2	1.33	- 0.65	0.22	0.01	- 2.15	0.24	0.6 3	1.67
03M06 2	19 9	0.87	- 0.17	0.27	0.29	- 1.17	0.24	1.0 7	2.58
3M017	19 6	1.08	- 0.84	0.45	0.77	- 1.89	0.19	1.1 2	2.05
3M032	18 2	1.08	- 0.38	1.43	2.22	- 1.45	0.68	1.7 5	2.20
3M033	19 9	1.35	- 0.67	- 0.19	- 0.89	- 1.46	- 0.14	0.3 7	2.22
3M045	15 9	0.85	0.06	1.00	1.39	- 1.32	0.46	1.5 7	2.20
3M063	21 0	1.20	- 0.46	0.86	1.38	- 1.78	0.66	1.4 1	2.02

Table H.44. IRT Item Analysis Results – Grade 4 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
04M003	166	1.44	-0.33	1.98	0.91	-0.70	0.97	1.22	2.32
04M008	176	1.10	-0.66	1.06	0.98	-1.32	0.63	1.55	2.47
04M011	171	0.75	-0.65	0.60	0.59	-2.09	0.19	1.21	2.31
04M017	253	0.78	-0.50	0.62	0.62	-1.79	0.50	1.18	2.50
04M023	246	1.18	-0.37	1.20	0.81	-1.00	0.54	1.08	2.58
04M039	172	1.18	-0.80	1.02	1.11	-1.03	0.41	1.39	2.46
04M048	176	0.97	-0.73	1.53	1.83	-1.08	0.72	1.86	2.83
04M050	171	1.12	-1.00	0.72	0.61	-2.33	0.60	1.15	2.24
04M058	166	1.04	-0.43	0.62	-0.07	-1.52	0.21	1.01	1.99
04M062	197	0.88	-0.90	0.95	0.78	-1.98	0.61	1.14	2.47
04M064	253	1.00	-0.66	0.27	0.20	-1.94	0.37	0.97	2.31
4M020	172	0.89	-0.49	1.39	1.91	-1.08	0.74	1.78	2.80
4M034	246	0.98	-0.24	0.61	0.63	-1.19	0.25	1.14	2.43
4M041	197	0.99	-0.93	0.53	0.56	-2.19	0.45	1.36	2.25

Table H.45. IRT Item Analysis Results – Grade 5 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
05M004	178	1.02	- 0.79	0.7 1	0.93	- 0.76	0.54	1.0 9	1.92
05M008	252	0.94	- 1.15	0.6 2	0.58	- 1.72	0.24	1.2 2	1.85
05M014	161	0.85	- 0.26	1.0 2	0.16	- 1.83	0.43	1.0 3	2.00
05M033	202	0.95	- 1.85	0.1 0	- 0.35	- 2.37	0.20	0.5 6	1.90
05M044	252	0.85	- 0.63	0.6 4	0.78	- 1.37	0.39	1.1 8	1.96
05M045	186	0.91	- 0.16	0.0 9	1.27	- 1.72	0.24	1.2 4	1.93
05M054	190	0.98	- 1.60	0.0 6	- 0.13	- 1.74	0.23	0.6 8	1.80
05M058	186	0.77	- 1.56	0.1 2	0.05	- 2.65	- 0.46	0.8 4	1.69
5M009	171	0.90	- 0.38	0.2 0	0.22	- 1.93	- 0.07	0.7 4	1.78
5M018	190	0.85	- 0.52	1.4 4	0.29	- 0.44	0.49	0.9 7	2.22
5M031	161	0.99	- 0.58	0.7 9	- 0.23	- 2.17	0.37	0.6 6	1.83
5M036	171	0.83	- 0.52	1.1 2	- 0.24	- 1.95	0.19	1.0 7	1.76
5M050	202	1.00	- 0.45	0.0 6	0.57	- 1.13	0.41	1.2 5	2.02
5M064	178	0.98	- 0.70	0.4 5	1.12	- 0.99	0.61	1.1 6	1.90

Table H.46. IRT Item Analysis Results – Grade 6 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
06M001	237	1.14	-0.84	0.99	0.88	-1.50	0.68	1.64	2.94
06M024	268	1.25	-0.84	0.28	0.46	-1.79	0.11	0.95	2.47
06M036	204	1.05	-0.36	0.66	1.41	-1.62	0.54	1.74	2.77
06M042	237	1.08	-0.42	1.11	1.41	-0.92	0.64	1.79	3.16
06M046	202	1.02	-0.64	-0.05	0.74	-2.00	0.02	1.06	2.52
06M052	204	0.91	-1.41	0.44	-0.43	-2.67	-0.17	1.01	2.38
06M059	202	0.96	-1.44	0.61	1.02	-2.51	0.34	1.11	2.70
06M060	182	1.01	-0.15	1.01	1.44	-0.81	0.68	1.49	2.67
06M065	268	1.14	-1.35	0.16	0.39	-2.36	0.06	1.01	2.40
06M101	217	0.88	-0.93	1.05	0.75	-1.93	0.57	1.42	2.54
6M007	192	1.17	-0.44	1.45	1.19	-1.37	0.59	1.82	2.59
6M017	217	1.05	-0.96	0.26	0.84	-2.23	0.33	1.53	2.37
6M040	182	1.04	-0.52	1.26	1.45	-1.05	0.69	1.77	2.59
6M050	192	1.08	-0.76	1.03	-0.20	-2.12	0.17	1.28	2.25

Table H.47. IRT Item Analysis Results – Grade 7 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
07M001	192	0.92	- 1.98	0.36	- 0.39	- 2.52	- 0.18	0.5 7	1.39
07M010	293	0.98	- 1.03	0.47	0.50	- 1.97	0.12	0.8 1	1.84
07M016	205	0.79	- 1.28	0.84	- 0.25	- 2.51	0.12	0.9 4	1.96
07M021	180	0.81	- 1.44	- 0.29	- 0.53	- 2.09	- 0.56	0.4 8	1.27
07M027	211	1.08	- 1.55	- 0.55	0.44	- 2.59	- 0.08	0.6 8	1.27
07M028	196	0.93	- 1.82	0.73	0.54	- 2.08	- 0.18	0.9 6	1.55
07M041	180	0.92	- 1.16	1.16	1.23	- 1.17	0.40	1.2 1	1.65
07M048	205	0.97	- 0.34	0.28	0.05	- 1.80	0.00	1.2 4	1.91
07M053	213	1.00	- 0.72	0.35	- 1.11	- 2.00	- 0.29	0.7 1	1.43
07M058	293	0.88	- 1.56	0.56	- 0.81	- 2.49	- 0.31	0.4 1	1.55
07M060	196	0.84	- 1.20	0.01	0.44	- 2.21	- 0.24	0.9 1	1.44
07M063	213	0.81	- 0.78	0.32	0.25	- 1.88	0.06	0.8 7	1.75
7M037	192	0.97	- 0.90	0.70	0.47	- 1.41	0.23	0.8 9	1.60
7M044	211	0.96	- 0.88	0.48	1.01	- 1.73	0.29	1.0 9	1.44

Table H.48. IRT Item Analysis Results – Grade 8 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
08M007	279	1.06	-0.68	0.82	0.34	-1.65	0.52	1.22	2.27
08M011	194	0.85	-0.60	0.37	0.52	-0.92	0.23	0.93	2.25
08M019	200	1.06	-0.28	0.77	-0.56	-1.65	-0.08	0.99	2.34
08M022	210	0.88	-1.16	0.94	-0.17	-2.06	0.18	0.88	2.05
08M024	201	1.03	-0.88	1.10	2.15	-0.90	0.41	1.83	2.65
08M031	226	0.86	-1.11	0.48	-0.06	-2.05	-0.15	0.94	1.94
08M041	201	0.89	-0.66	0.40	0.82	-1.37	0.20	1.20	2.32
08M048	210	0.81	-1.47	0.16	0.27	-2.81	0.18	0.90	2.04
08M051	194	1.17	-0.39	0.77	0.30	-0.28	0.10	1.41	2.11
08M054	220	0.72	-1.14	1.39	0.28	-2.16	0.33	1.15	2.76
08M059	279	0.98	-1.51	0.60	0.41	-2.30	0.22	1.23	2.24
8M013	226	1.28	-0.41	0.82	1.99	-1.09	0.60	1.62	2.22
8M034	220	1.01	-0.68	0.52	1.12	-1.62	0.15	1.31	2.83
8M063	200	1.11	-1.54	-0.13	-0.24	-2.30	-0.63	0.48	2.30

Table H.49. IRT Item Analysis Results – Grade 11 Mathematics									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
11M00 4	26 0	0.86	- 1.07	0.11	- 0.25	- 2.79	0.03	0.9 2	1.70
11M00 6	15 0	0.84	- 0.84	0.05	1.33	- 2.02	- 0.04	1.2 3	2.32
11M01 1	14 8	0.90	- 1.61	- 0.09	0.42	- 2.81	- 0.20	1.0 6	1.74
11M01 4	17 0	0.69	- 0.49	0.41	- 1.04	- 2.40	0.02	0.7 0	1.75
11M01 5	13 3	1.13	- 0.17	- 0.32	- 0.39	- 0.92	0.26	0.9 5	1.68
11M03 2	10 6	1.03	- 0.48	0.56	0.46	- 2.06	0.23	1.2 0	1.38
11M03 8	18 1	0.96	- 0.98	- 0.25	0.33	- 2.11	0.02	0.8 5	1.38
11M03 9	10 6	0.88	- 1.48	0.70	- 0.77	- 3.22	0.00	0.4 4	1.31
11M04 7	18 1	0.91	- 0.98	- 0.09	0.53	- 2.08	0.10	0.9 0	1.44
11M05 0	17 0	1.06	- 2.51	- 0.31	- 0.43	- 3.37	- 0.59	0.2 1	1.76
11M05 5	15 0	0.86	- 0.43	- 0.49	0.08	- 2.46	0.01	1.0 0	1.90
11M06 3	26 0	1.00	- 1.15	0.73	1.59	- 2.26	0.63	1.3 7	2.10
11M06 4	14 8	0.74	- 1.13	0.32	0.06	- 2.35	- 0.16	0.9 9	1.78
11M10 1	13 3	1.21	0.39	- 0.24	0.74	- 0.39	0.65	1.2 8	1.84

Table H.50. IRT Item Analysis Results - Grade 3 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
03R003	211	0.99	-0.31	0.84	1.79	-2.18	0.77	1.84	2.62
03R010	182	1.10	-0.48	1.07	1.38	-2.01	0.79	1.64	2.33
03R012	196	1.09	-0.66	1.00	0.48	-1.48	0.13	1.42	2.41
03R020	159	1.22	-1.68	0.42	-0.11	-2.31	-0.28	1.17	2.12
03R028	155	0.96	0.45	0.85	1.89	-0.96	0.73	1.89	2.96
03R034	151	1.31	-1.18	1.04	1.14	-1.33	0.82	1.56	2.40
03R041	199	0.89	-0.87	0.38	1.07	-1.86	0.16	1.72	2.93
03R042	151	1.03	-0.41	1.13	1.92	-0.76	0.73	2.04	2.61
03R047	159	1.09	0.05	1.59	2.09	-0.93	0.93	1.98	2.87
03R050	196	1.29	0.30	1.24	1.58	-0.95	0.94	1.83	2.64
03R100	211	0.90	-0.27	0.63	1.02	-2.30	0.48	1.46	2.53
03R101	182	1.33	-0.29	1.50	2.51	-1.44	0.96	2.06	2.38
3R026	199	1.08	-0.53	0.26	1.01	-1.75	0.58	1.53	2.91
3R038	155	0.97	-0.88	0.43	1.54	-2.18	0.51	1.64	2.74

Table H.51. IRT Item Analysis Results - Grade 4 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
04R00 2	25 3	1.17	- 0.74	0.67	1.0 9	- 1.83	0.81	1.5 5	2.30
04R00 9	16 6	0.93	- 0.39	0.19	0.2 9	- 1.96	0.25	0.9 1	2.21
04R01 4	19 7	0.79	- 1.36	- 0.10	0.3 8	- 2.91	- 0.10	0.9 2	2.37
04R01 9	24 6	1.09	- 0.88	1.16	1.2 3	- 1.36	0.49	1.5 7	2.30
04R02 3	16 6	1.09	- 0.35	1.38	0.8 3	- 1.42	0.81	1.4 9	2.42
04R02 4	19 7	0.94	- 0.79	0.93	0.5 6	- 2.23	0.68	1.3 5	2.44
04R02 6	24 6	0.94	0.05	1.03	0.6 7	- 1.03	0.49	1.3 6	2.29
04R03 3	17 1	0.85	- 0.31	0.81	1.7 4	- 1.73	0.56	1.7 6	2.69
04R03 8	17 1	1.18	- 0.53	1.09	1.6 3	- 1.96	1.01	1.8 4	2.52
04R04 3	17 2	0.93	- 1.29	1.18	0.9 5	- 1.59	0.41	1.5 8	2.32
04R04 6	17 6	0.93	0.00	1.38	1.3 8	- 0.80	0.81	1.7 3	2.56
04R04 8	25 3	1.11	- 0.34	0.84	1.1 1	- 1.47	0.76	1.7 0	2.35
04R05 1	17 6	0.87	0.09	1.60	1.6 0	- 0.78	0.96	1.9 1	2.64
04R05 3	17 2	0.98	- 0.23	1.81	2.0 3	- 0.70	1.01	1.8 8	2.72

Table H.52. IRT Item Analysis Results - Grade 5 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
05R00 1	25 2	0.97	- 1.09	0.9 0	0.93	- 1.85	0.4 1	1.5 9	2.43
05R00 4	18 9	1.24	- 0.59	1.0 1	0.95	- 0.36	0.8 1	1.5 8	2.21
05R00 8	18 9	0.95	- 0.45	1.4 3	1.16	- 0.38	0.8 6	1.6 0	2.45
05R01 5	16 1	1.25	- 0.45	1.9 8	0.43	- 1.30	0.9 1	1.6 3	2.43
05R02 1	16 1	1.07	- 0.38	0.7 1	1.69	- 1.41	0.7 1	1.4 4	2.67
05R02 7	17 1	1.05	- 0.87	1.1 5	1.06	- 2.37	0.7 4	1.4 7	2.23
05R02 9	20 2	1.06	0.09	1.1 2	1.36	- 0.45	1.0 7	1.6 2	2.70
05R03 3	17 8	1.12	- 0.60	0.7 2	1.25	- 0.98	1.0 1	1.2 9	2.25
05R03 6	18 6	0.85	- 1.15	1.2 9	0.99	- 2.40	0.2 2	1.6 2	2.49
05R04 7	20 2	1.01	- 1.42	0.8 7	- 0.30	- 1.42	0.2 1	0.9 0	2.32
05R05 2	17 8	1.02	- 0.48	0.5 8	0.99	- 0.90	0.4 9	1.5 7	2.15
05R05 5	18 6	0.92	- 0.81	1.3 3	1.73	- 2.05	0.4 2	1.8 6	2.65
05R06 4	17 1	0.99	- 0.47	1.2 4	1.70	- 1.87	0.8 2	1.6 1	2.44
5R018	25 2	0.92	- 0.31	0.7 9	1.48	- 1.26	0.5 8	1.6 5	2.62

Table H.53. IRT Item Analysis Results - Grade 6 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
06R001	268	0.97	- 1.07	0.5 4	0.91	- 1.66	0.27	1.1 4	2.18
06R009	192	1.25	- 1.05	1.1 7	1.57	- 1.81	0.66	1.5 8	2.16
06R019	237	1.09	- 1.56	0.1 8	0.78	- 1.87	0.12	1.2 1	2.49
06R023	217	0.84	- 0.64	0.4 4	0.40	- 1.75	0.46	1.0 9	2.21
06R027	201	0.81	- 0.93	0.4 0	1.08	- 2.09	0.37	1.1 6	2.26
06R032	268	1.00	- 0.96	1.0 2	1.19	- 1.56	0.57	1.3 6	2.29
06R034	192	0.99	- 1.45	1.0 3	0.82	- 2.42	0.29	1.6 1	2.04
06R038	204	1.03	- 0.47	0.6 4	1.31	- 1.65	0.68	1.5 7	2.30
06R045	182	0.94	- 0.50	0.4 8	0.28	- 1.56	0.43	0.8 2	2.08
06R049	181	0.89	- 0.97	0.4 8	0.84	- 1.93	0.33	1.2 9	2.12
06R057	217	0.97	- 1.22	1.2 2	0.44	- 1.54	0.42	1.3 2	2.29
06R061	237	1.17	- 0.64	0.9 1	2.14	- 1.30	0.94	1.9 6	2.75
06R100	204	1.06	- 0.07	0.1 7	0.92	- 1.59	0.54	1.3 8	2.21
6R010	201	1.01	- 0.95	0.0 3	- 0.32	- 2.05	- 0.52	0.8 3	1.89

Table H.54. IRT Item Analysis Results - Grade 7 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
07R001	294	1.06	- 0.85	0.7 3	1.2 3	- 2.02	0.59	1.1 8	2.77
07R009	192	1.05	- 1.41	0.5 6	0.4 2	- 1.68	0.29	0.9 5	2.27
07R015	205	1.09	- 0.71	0.6 4	0.2 2	- 1.54	0.01	1.1 6	2.76
07R016	294	0.92	- 0.61	0.8 9	1.2 3	- 1.70	0.29	1.5 3	2.80
07R019	180	0.92	- 0.91	0.8 5	1.0 6	- 1.95	0.62	1.7 6	2.33
07R026	211	1.21	- 0.95	0.3 5	0.7 4	- 1.77	0.36	1.4 9	2.07
07R030	180	0.97	- 0.92	0.9 6	1.7 6	- 1.70	0.71	1.8 6	2.50
07R033	192	1.27	- 1.45	0.5 7	1.8 2	- 1.82	0.74	1.5 5	2.40
07R040	205	1.44	- 0.80	0.0 6	1.5 4	- 1.57	0.28	1.7 1	2.85
07R044	197	1.17	- 1.84	0.2 0	1.0 0	- 2.11	- 0.08	0.9 6	2.47
07R050	213	1.28	- 0.70	0.7 4	1.1 9	- 1.36	0.55	1.6 3	2.37
07R055	211	1.05	- 0.06	0.5 5	1.4 2	- 1.07	0.63	1.5 5	2.37
07R058	213	1.21	- 0.42	1.0 1	1.6 8	- 1.25	0.94	1.6 7	2.57
7R028	197	1.24	- 1.27	0.4 8	1.0 8	- 1.91	0.17	1.4 1	2.41

Table H.55. IRT Item Analysis Results - Grade 8 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
08R004	278	1.21	- 0.59	1.1 7	1.98	- 1.42	0.90	2.1 3	2.82
08R009	201	0.84	- 0.96	0.1 5	0.31	- 2.70	- 0.01	1.5 0	2.59
08R013	210	0.85	- 0.61	0.7 9	1.42	- 1.61	0.62	1.6 6	2.84
08R018	221	0.89	- 0.56	0.8 5	0.76	- 1.68	0.35	1.4 4	2.88
08R021	201	0.88	- 0.36	1.2 9	1.93	- 1.76	1.04	1.9 6	3.15
08R026	210	0.96	- 0.82	0.8 2	0.22	- 1.66	0.24	1.3 0	2.54
08R027	278	0.75	- 1.49	0.3 4	- 0.14	- 2.63	- 0.18	0.9 6	2.44
08R030	201	1.02	0.13	1.4 2	1.53	- 0.60	0.88	1.9 5	3.03
08R035	221	0.88	- 1.30	0.3 5	- 0.22	- 2.49	0.05	0.7 3	2.59
08R043	201	1.15	- 0.07	0.4 7	1.02	- 1.04	0.52	1.8 5	2.67
08R049	194	1.14	- 1.06	1.3 0	1.65	- 0.51	0.63	1.8 0	2.67
08R061	194	1.33	- 0.74	1.1 6	1.08	- 0.73	0.91	1.4 8	2.47
08R100	226	0.98	- 0.78	0.9 5	1.31	- 1.70	0.45	1.8 2	2.66
8R052	226	1.09	- 0.96	1.3 0	1.33	- 1.59	0.54	1.7 4	2.74

Table H.56. IRT Item Analysis Results - Grade 11 Reading									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
11R00 2	26 2	0.84	0.05	0.22	0.8 1	- 1.72	0.39	1.7 3	3.23
11R01 1	17 0	1.02	0.21	0.92	0.7 5	- 1.33	0.64	1.7 9	2.98
11R02 1	17 0	1.52	0.30	0.91	2.2 0	- 0.85	1.15	2.2 8	3.12
11R02 3	13 4	1.21	0.55	1.01	2.0 2	0.06	1.69	2.0 3	3.52
11R02 5	13 4	1.14	0.20	0.70	0.0 7	- 0.70	0.77	1.7 1	3.09
11R02 6	18 1	1.23	- 0.55	0.50	0.7 6	- 1.54	0.00	1.7 9	2.77
11R03 0	18 1	1.31	- 0.75	0.55	0.8 9	- 2.05	0.69	1.6 9	2.78
11R03 3	10 6	1.03	- 1.23	0.34	1.4 5	- 3.18	0.48	1.3 4	2.79
11R04 1	14 8	0.94	- 0.52	0.16	1.3 9	- 1.89	0.38	1.8 2	3.17
11R04 5	14 8	0.98	- 1.14	0.46	0.0 8	- 1.92	0.03	0.4 5	3.04
11R04 7	15 0	0.86	- 1.23	- 0.04	0.3 2	- 3.39	- 0.02	1.2 1	2.78
11R05 0	15 0	0.78	- 1.97	0.43	0.5 2	- 3.44	- 0.02	1.0 6	2.90
11R05 4	10 6	1.07	- 0.93	- 0.03	0.9 4	- 3.12	0.18	1.1 8	2.61
11R10 0	26 2	1.22	0.18	0.53	2.9 1	- 0.98	0.70	2.4 6	3.64

Table H.57. IRT Item Analysis Results - Grade 4 Science									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
04S00 2	25 3	0.89	- 0.67	0.46	0.62	- 1.93	0.21	1.2 6	2.60
04S00 6	25 3	1.22	- 0.24	1.00	1.53	- 1.34	0.84	1.6 5	2.81
04S01 0	16 6	1.55	- 1.12	0.70	0.37	- 2.09	0.78	1.0 5	2.35
04S01 7	19 6	1.24	- 0.15	1.49	2.02	- 1.17	1.24	2.0 1	3.09
04S02 3	24 4	1.28	- 0.21	1.03	1.41	- 1.23	0.69	1.7 2	2.63
04S02 7	19 6	0.85	- 0.24	0.80	0.86	- 1.76	0.58	1.5 7	2.82
04S02 9	24 4	1.15	- 0.65	1.08	- 0.08	- 1.65	0.24	1.2 0	2.36
04S03 4	17 1	1.34	- 1.03	2.20	2.18	- 1.94	1.35	1.8 8	2.82
04S03 6	17 1	1.11	- 0.35	1.04	1.63	- 1.91	0.89	1.8 3	2.59
04S04 5	17 2	1.10	- 1.46	- 0.10	0.33	- 2.10	- 0.45	1.2 1	2.26
04S04 8	17 5	1.04	- 0.19	- 0.09	0.23	- 1.53	- 0.48	1.0 3	2.44
04S05 3	16 6	0.77	- 0.64	0.60	- 0.25	- 2.05	- 0.11	1.1 4	2.36
04S05 4	17 2	1.33	- 0.21	1.32	2.26	- 0.78	0.96	1.6 7	2.95
04S05 8	17 5	1.23	- 1.02	0.92	1.56	- 1.77	0.61	1.8 6	2.63

Table H.58. IRT Item Analysis Results - Grade 7 Science									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
07S007	292	0.72	-0.85	0.77	0.22	-2.35	0.23	1.19	2.61
07S016	180	0.86	-1.18	0.86	-1.14	-2.08	-0.22	1.21	2.25
07S020	211	1.05	-0.63	0.27	1.30	-1.96	0.71	1.64	2.48
07S023	191	0.73	-0.82	0.81	0.66	-1.74	0.36	1.54	2.68
07S028	205	1.53	-1.19	0.94	1.45	-2.18	1.04	1.87	2.79
07S031	196	1.27	-0.33	1.57	1.37	-1.21	0.90	1.90	2.57
07S032	213	1.17	0.12	1.18	1.13	-1.12	0.99	1.69	2.67
07S035	180	1.01	-0.67	1.57	2.30	-0.87	0.94	1.83	3.18
07S039	211	1.30	-0.85	0.71	1.40	-1.53	0.67	1.78	2.49
07S041	191	1.23	-1.05	0.90	2.37	-1.05	0.62	1.96	2.98
07S047	205	1.09	-0.37	1.69	1.56	-1.04	0.87	1.94	3.12
07S049	196	1.24	-0.42	1.62	2.21	-1.36	1.20	2.05	2.68
07S054	292	0.96	-0.78	0.31	0.90	-2.25	0.34	1.27	2.69
07S056	213	0.97	-0.40	0.43	1.15	-1.57	0.35	1.40	2.64

Table H.59. IRT Item Analysis Results - Grade 11 Science									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
11S001	260	0.82	- 1.28	0.66	1.05	- 2.26	0.29	1.47	2.60
11S009	170	0.89	0.14	0.82	1.32	- 1.35	0.65	1.82	2.87
11S010	260	0.98	- 1.51	0.55	1.28	- 2.62	0.42	1.71	2.52
11S015	133	1.05	0.91	0.57	1.57	0.03	1.11	1.75	2.94
11S018	170	1.04	- 0.75	0.55	0.80	- 1.57	- 0.14	1.66	2.65
11S019	148	0.93	- 0.57	1.23	0.84	- 1.43	0.72	1.41	2.57
11S022	149	0.84	- 0.25	0.87	0.31	- 1.60	0.70	1.27	2.18
11S032	181	1.01	- 1.80	0.47	0.97	- 2.22	0.18	1.17	2.17
11S034	181	0.88	- 0.73	0.40	0.55	- 1.84	0.35	1.15	2.09
11S041	106	0.82	- 1.20	0.41	0.22	- 2.88	- 0.13	1.20	1.96
11S046	106	0.81	- 1.34	0.61	0.40	- 2.99	0.11	1.15	2.05
11S050	149	0.94	0.05	- 0.59	0.17	- 2.20	0.46	1.38	1.95
11S101	133	0.92	0.66	0.73	0.86	- 0.27	1.06	1.60	2.78
11S102	148	0.79	- 1.51	0.64	0.36	- 2.43	0.22	1.15	2.38

Table H.60. IRT Item Analysis Results - Grade 5 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
5W019	186	0.83	- 0.69	0.2 4	1.12	- 2.22	- 0.05	1.1 5	2.56
5W035	202	0.83	- 0.55	0.5 1	0.92	- 1.83	0.46	1.4 2	2.73
5W101	161	0.81	- 0.64	1.5 5	0.95	- 2.02	0.60	1.3 7	2.91
5W115	170	0.92	- 0.57	1.1 6	0.93	- 2.09	0.67	1.1 8	2.78
5W131	190	0.99	- 1.10	0.1 3	0.77	- 1.48	0.34	1.0 1	2.68
5W136	177	0.79	- 0.60	0.2 7	0.77	- 1.46	0.24	1.2 8	2.26
5W202	252	1.06	- 1.85	0.4 8	- 0.35	- 2.67	- 0.16	1.0 0	2.10

Table H.61. IRT Item Analysis Results - Grade 6 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
6W102	268	1.08	- 1.94	1.00	1.74	- 2.51	0.11	1.4 9	2.46
6W103	202	1.10	- 1.08	0.09	0.89	- 2.14	0.06	0.7 2	2.33
6W104	192	1.15	- 1.19	0.71	0.94	- 2.33	0.20	1.3 3	2.01
6W106	204	0.65	- 1.07	- 0.03	- 0.43	- 2.82	- 0.45	0.9 0	1.92
6W108	217	1.07	- 1.33	0.45	- 0.35	- 2.19	- 0.06	0.7 9	1.79
6W109	236	1.18	- 1.58	- 0.08	0.04	- 2.72	- 0.12	0.8 2	2.21
6W119	182	1.05	- 2.04	0.02	0.71	- 2.78	- 0.24	0.8 2	2.05

Table H.62. IRT Item Analysis Results - Grade 8 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
8W059	200	1.08	- 0.78	1.23	1.56	- 1.61	0.38	1.58	3.07
8W101	278	0.96	- 1.31	0.27	0.20	- 2.67	- 0.06	1.04	2.61
8W102	200	0.78	- 1.45	- 0.05	- 0.05	- 2.93	- 0.20	0.58	2.66
8W103	210	1.08	- 0.42	- 0.17	1.24	- 2.20	- 0.31	1.48	2.73
8W107	221	0.88	- 1.40	1.25	1.18	- 1.95	0.04	1.51	3.21
8W111	194	1.20	- 1.45	1.01	1.59	- 1.73	0.74	1.38	2.81
8W113	226	0.99	- 1.69	0.39	0.44	- 2.64	- 0.33	0.99	2.41

Table H.63. IRT Item Analysis Results - Grade 11 Writing									
			Step Values			Average Measure			
Item	N	In-Fit	2	3	4	1	2	3	4
9W103	170	0.93	- 0.18	0.99	1.30	- 1.79	0.52	2.00	2.91
9W202	261	0.68	- 0.77	0.03	0.20	- 2.49	- 0.26	0.95	2.74
9W205	134	1.07	- 0.15	- 0.05	0.28	- 1.04	0.06	1.38	2.64
9W208	181	0.81	- 1.13	0.14	- 0.33	- 2.15	- 0.12	0.60	2.10
9W211	106	1.06	- 0.69	- 0.56	0.60	- 3.32	0.32	0.83	2.39
9W213	150	1.03	- 3.24	- 0.56	0.92	- 4.38	- 0.65	1.17	2.79
9W216	148	0.65	- 1.09	- 0.79	- 0.08	- 2.63	- 0.25	0.49	2.40