

**TECHNICAL REPORT
NEW JERSEY
HIGH SCHOOL PROFICIENCY
ASSESSMENT**

Spring 2015
Measurement Incorporated™

**February 2016
PTM: 1510.67**

**Copyright © 2016 by New Jersey Department of Education
All rights reserved**

STATE BOARD OF EDUCATION

- MARK W. BIEDRON Hunterdon
President
- JOSEPH FISICARO Burlington
Vice President
- ARCELIO APONTE.....Middlesex
- RONALD K. BUTCHER Gloucester
- JACK FORNARO..... Warren
- EDITHE FULTON Ocean
- ERNEST P. LEPORE Hudson
- ANDREW J. MULVIHILL Sussex
- J. PETER SIMON Morris
- DOROTHY STRICKLAND..... Essex

DAVID C. HESPE, Commissioner of Education

Table of Contents

Table of Contents	iii
Part 1: Introduction	5
1.1 Description of Assessment	5
1.2 Purpose of Assessment	6
1.3 Organizational Support.....	7
Part 2: Test Development.....	9
2.1 Test Specifications	9
2.2 Development of Test Items	11
2.3 Item Review Process	14
2.4 Item Use	15
2.5 Quality Control for Test Construction.....	16
Part 3: Test Administration.....	17
3.1 Participation	17
3.2 Test Security Procedures	17
3.3 Test Administration Procedures	19
3.4 Test Accommodations	22
Part 4: Scoring.....	32
4.1 MC items	32
4.2 OE items	32
4.3 Quality Control Procedures in Data Preparation	36
Part 5: Item-Level Statistics.....	37
5.1 Classical Item Statistics.....	37
5.2 Speededness	40
5.3 Intercorrelations.....	40
Part 6: Standard Setting	42
Part 7: Scaling and Equating.....	44
7.1 Scaling.....	44
7.2 Equating	44
Part 8: Test Statistics.....	56
8.1 Summary Statistics	56
8.2 Classical Reliability Estimates of the Test Scores	56
8.3 Reliability of Performance Classifications.....	60
8.4 Conditional Estimate of Error at Each Cut Score.....	61
8.5 Rater Reliability	62
Part 9: Validity.....	64
Part 10: Reporting.....	65
10.1 Cycle I Reports.....	65
10.2 Cycle II Reports.....	68
10.3 State Summary Report (Reporting Rules).....	68
10.4 Interpreting Reports.....	73
10.5 Quality Control in Reporting.....	73
References.....	77
Appendix A: Test Specifications from the February 1998 Report	78
Appendix B: Raw to Scale Score Conversions and Frequency Distributions	83
Appendix C: Score Calculation Figures and Scoring Rubrics.....	88
Appendix D: Standard Error Plots	95
Appendix E: Standard Setting Executive Summary	98

Tables and Figures

Table 1.1.1: 2015 Retesting Student Population State Level Performance	6
Table 1.3.1: Contributors and Their Roles/Responsibilities	7
Table 2.1.1: 2015 March NJ HSPA Total Points By Content Area	11
Table 3.4.1: Special Education (Setting Accommodation)	28
Table 3.4.2: Special Education (Scheduling Accommodations)	29
Table 3.4.3: Special Education (Test Materials Modifications)	30
Table 3.4.4: Special Education (Test Procedure Modifications)	31
Table 5.1.1: Item Difficulty and Discrimination Summary Statistics for Multiple-Choice and Open-Ended items by Content Area and Cluster (Standard)	38
Table 5.1.2: Frequency Distribution for Multiple-Choice P-values and Discrimination Indices by Content Area and Cluster (Standard)	39
Table 5.2.1: NJ HSPA Percentage of Students Omitting the Last Item of Each Part	40
Table 5.3.1: NJ HSPA Intercorrelations Among Major Content Clusters (Standards) and Item Types	41
Table 5.3.2: NJ HSPA Intercorrelations Among Content Areas and Clusters (Standards)	41
Table 6.1.1: Target HSPA Cut Scores and Rasch thetas from Spring 2002 Standard Setting	42
Table 6.1.2: Obtained HSPA Math Spring Cut Scores	42
Table 6.1.3: Obtained HSPA LAL Spring Cut Scores	42
Table 6.1.4: Obtained Math Rasch thetas for the raw Cut Scores in Table 6.1.2	43
Table 6.1.5: Obtained LAL Rasch thetas for the raw Cut Scores in Table 6.1.3	43
Table 7.2.1: Sample Demographics for LAL 2002, 2011-2015 General Population	45
Table 7.2.3: Language Arts Literacy Infit/Outfit Summary Statistics	46
Table 7.2.4: Sample Demographics for MATH 2002, 2011-2015 General Population	47
Table 7.2.6: Mathematics Infit/Outfit Summary Statistics	48
Figure 7.2.1 LAL Item Map	49
Figure 7.2.2 Math Item Map	50
Table 7.2.7: Language Arts Literacy Open-Ended Score Distribution	54
Table 7.2.8: Mathematics Open-Ended Score Distribution	55
Table 8.1.1: NJ HSPA Raw Score Means and Standard Deviations	56
Table 8.1.2: 2015 NJ HSPA Means and Standard Deviations of Students' Raw Score and Percent Correct by Cluster (Standard)	56
Table 8.2.1: 2012 NJ HSPA Reliability Estimates and Standard Errors of Measurement for Content Areas and Clusters (Standards)	57
Table 8.2.2: NJ HSPA Reliability Estimates and Standard Errors of Measurement for Content Areas and Clusters (Standards) for Subgroups in Language Arts	58
Table 8.2.3: NJ HSPA Reliability Estimates and Standard Errors of Measurement for Content Areas and Clusters (Standards) for Subgroups in Mathematics	59
Table 8.3.1: NJ HSPA KR-20	60
Table 8.3.2: NJ HSPA Language Arts Literacy Decision Consistency	60
Table 8.3.3: NJ HSPA Mathematics Decision Consistency	61
Table 8.4.1: NJ HSPA Conditional Estimate of Error at Each Cut-Score	61
Table 8.5.1: NJ HSPA Spring 2015 Reader Agreement Statistics	63
Table B.1: Language Arts Literacy Raw to Scale Score Conversion and Frequency Distribution	84
Table B.2: Mathematics Raw to Scale Score Conversion and Frequency Distribution	86
Table C.1: Score Calculation Figures	89
Table C.2: Open-Ended Mathematics Scoring Rubric	92
Table C.3: New Jersey Registered Holistic Scoring Rubric	93
Table C.4: Open-Ended Scoring Rubric (Reading)	94

Part 1: Introduction

1.1 Description of Assessment

In 1975, the New Jersey Legislature passed the Public School Education Act “to provide to all children in New Jersey, regardless of socioeconomic status or geographic location, the educational opportunity which will prepare them to function politically, economically and socially in a democratic society.” An amendment to that act was signed in 1976 which established uniform standards of minimum achievement in basic communication and computation skills. This amendment is the legal basis for the use of a test as a graduation requirement in the state of New Jersey.

Beginning in 1981–1982, ninth-grade students were required to pass the Minimum Basic Skills Test in Reading and Mathematics as one of the requirements for a high school diploma. Students who did not pass both sections of the test had to be retested on those sections not passed.

In 1983, a more difficult test in Reading, Mathematics, and Writing was adopted, the Grade 9 High School Proficiency Test (HSPT9), to measure the basic skills achievements of ninth-grade students. The first due-notice administration of the HSPT9 occurred in the 1983–1984 school year. The first year the test was administered as a graduation requirement was 1985–1986.

In 1988, the New Jersey Legislature passed a law which moved the High School Proficiency Test from the ninth grade to the eleventh grade. The Grade 11 High School Proficiency Test (HSPT11), a rigorous test of essential skills in Reading, Mathematics, and Writing, served as a graduation requirement for all public school students in New Jersey who entered the ninth grade on or after September 1, 1991. Three years of due-notice testing were conducted to allow school districts time to modify curricula and prepare students for the graduation test.

In May 1996, the New Jersey State Board of Education adopted Core Curriculum Content Standards to describe what all students should know and be able to do at the end of fourth grade, eighth grade, and upon completion of a New Jersey public school education. The Core Curriculum Content Standards delineate New Jersey’s expectations for student learning. All New Jersey school districts are required to organize instruction and design curricula so that virtually all students achieve the new content standards. The Core Curriculum Content Standards define the state’s high school graduation requirements and its testing program to measure benchmark achievements toward those requirements in grades 4, 8, and 11.

The Elementary School Proficiency Assessment (ESPA), which was administered to fourth-graders, was designed from its inception in 1997 to align with the content standards, as is the New Jersey Assessment of Skills and Knowledge (ASK), which replaced the ESPA and was expanded to include third-graders. As a result of the No Child Left Behind Act, the New Jersey Assessment of Skills and Knowledge was expanded to the fifth, sixth, and seventh grades in 2006. The Grade Eight Proficiency Assessment (GEPA), which replaced the Grade 8 Early Warning Test (EWT), is aligned with the content standards. The GEPA should be used for placement purposes and program planning so that students are provided with the appropriate instruction to enable them to go on and pass the state’s graduation test. The GEPA was replaced by NJ ASK 8, which served the same purpose as the GEPA. The High School Proficiency

Assessment (HSPA) is also aligned with the content standards and has replaced the HSPT11 as the state’s graduation test. Beginning in March 2002, the HSPA became the graduation test of record for all students who entered the eleventh grade for the first time on or after September 1, 2001. All assessment results are also used to satisfy federal requirements under the No Child Left Behind Act. In 2015 the state of New Jersey moved their assessments of ELA and math to the Partnership for Assessment of Readiness for College and Careers (PARCC). Thus, the population of examinees taking the 2015 HSPA assessment consisted of students who were retesting after previously scoring below the state’s graduation testing requirement.

The New Jersey HSPA scores are reported as scale scores and performance levels in each of the content areas. Students with scale scores ranging from 100 to 199 are classified as Partially Proficient, from 200 to 249 are classified as Proficient, and from 250 to 300 are classified as Advanced Proficient. State level performance of the retesting students is shown in Table 1.1.1. The vast majority of New Jersey high school students took the PARCC examination in 2015, as opposed to HSPA; thus, there are a reduced number of students in Table 1.1.1 in comparison to previous HSPA administrations.

Table 1.1.1: 2015 Retesting Student Population State Level Performance

Test Section	Number of Students	Proficiency Levels						Scale Score Mean
		Partially Proficient (100 – 199)		Proficient (200 – 249)		Advanced Proficient (250 – 300)		
		N	%	N	%	N	%	
LAL	3,599	2544	0.71	1011	0.28	44	0.01	185.50
Math	11,048	9153	0.83	1855	0.17	40	0.00	181.11

1.2 Purpose of Assessment

The initial purpose of the NJ HSPA was to provide annual feedback to students, teachers, parents, schools, districts, the State, and the general public in regard to the achievement of individual students and the success of educational programs in relation to the state curriculum. NJ HSPA also served the secondary purpose of providing diagnostic information to teachers regarding strengths and weaknesses of students within the three broad areas assessed. The dual purpose of the tests is reflected in the score reports, which contain not only total scores and designations of achievement level, but also cluster-level scores for students and programs. No scores for the individual skills that make up the clusters (see the separate content area sections of this document, Appendix A) are reported at any level, which include student/parent, classroom, building, district, and state.

In addition to assessing student knowledge of subject-specific content, the NJ HSPA tests also assess student critical thinking skills. Critical thinking skills are a primary focus of each of the tests. These skills are assessed through multiple-choice (MC), open-ended (OE), and writing prompt (W) items. The blueprints included in the subject-matter sections (Appendix A) of this

document reflect the crossing of content with process.

The purpose of the NJ HSPA tests is to accurately measure students’ achievement in Language Arts Literacy and Mathematics. To achieve this purpose, school personnel administering the tests play a crucial role. Test administrators can minimize problems that interfere with testing students by addressing the following guidelines:

- Maintain the security of all testing materials before, during, and after testing;
- Administer the tests in a manner consistent with established testing procedures;
- Establish testing conditions that prevent opportunities for students to engage in irregular behaviors (intentional or unintentional); and
- Communicate with the district NJ HSPA coordinator if questions arise.

1.3 Organizational Support

Measurement Incorporated, the contractor selected by New Jersey to develop, administer, score, and report on the NJ HSPA, understands the need for proactive communication among all parties involved in a successful assessment program. Given the nature of a project as large and diverse as the NJ HSPA program, numerous individuals, departments, and the contractor fill specific roles and responsibilities. Table 1.3.1 shows the major contributors to the NJ HSPA and their general responsibilities.

Table 1.3.1: Contributors and Their Roles/Responsibilities

Contributor	Role/Responsibility
NJ DOE	Oversight of the entire NJ HSPA program and content standards
NJ HSPA Program Coordinator	Daily management of the NJ HSPA program
NJ HSPA Content Specialist	Content-specific item development and testing issues
NJ HSPA Item Review Committees	Review and approval of all test items and passages
Technical Advisory Committee	Advises and assists on technical guidelines of the NJ HSPA program.
Educators	Provide feedback on all aspects of the NJ HSPA program. Administer the tests to students. Serve on content item review committees.
Measurement Incorporated (MI)	Develop and provide all products and materials relating to the testing, standard setting, scoring, and reporting for the NJ HSPA program.
MI HSPA Project Director	Daily management of the NJ HSPA program at MI

It is this need for coordination among various entities within the NJ DOE and Measurement Incorporated that necessitates the use of project management and production schedules, progress reports, conference calls, management and planning meetings, and meeting minutes.

Additionally, for each test administration a District Test Coordinator Manual and Test Examiner Manual is provided to the appropriate administrators and teachers involved in the testing process.

Part 2: Test Development

2.1 Test Specifications

Development of HSPA Test Forms

The HSPA consists of tests in Language Arts Literacy and Mathematics. Language Arts Literacy and Mathematics were given twice a year, through 2014. In 2015, the HSPA was administered only in the spring. Each test has a similar base test design, although the spring test includes embedded field test items distributed over multiple forms. The position of these items does not vary from year to year. Definitions of specific terminology (Narrative, Persuasive, Macros, Clusters, etc.) in these blueprints may be found in Appendix A of this document. These definitions were taken from the February 1998 Test Specifications document for the LAL test.

Language Arts Literacy (LAL)

The Language Arts Literacy test consists of two passages, each with 10 multiple-choice and two open-ended questions, and two writing tasks. One of the passages is a persuasive passage and the other is a narrative passage. One of the writing tasks, the Expository Prompt, asks students to write an essay which explains a subject to the reader, and in order to do so, the writer must establish the main idea and present it with supporting facts and/or pertinent information. The other writing task, the Persuasive Writing Prompt, asks students to write an essay intended to persuade a reader of a particular viewpoint. A general topic is provided for the essay.

The total possible point for Language Arts Literacy is 54 points, with scores changing in half-point increments. This includes

1 point for each of 20 multiple-choice items = 20

4 points for each of 4 open-ended items (two readers' scores averaged) = 16

6 points for a Expository Prompt (two readers' scores averaged) = 6

12 points for a Writing Prompt (two readers' scores summed) = 12

The content of the test consists of both Working with Text and Analyzing Text skills. There are no formal requirements about how many points or questions of each kind, although student scores for both are reported. Items with passages are assigned to one of these skills and scores are reported for these categories. The two writing tasks are not assigned to these categories. Tests are assembled using p -values and means of items from field-testing and are equated using a Rasch model with theta for the appropriate cut scores.

Mathematics (Math)

The Mathematics test consists of all independent items. Each test has 30 multiple-choice and 6 open-ended questions. The test measures content from four mathematics content standards:

- I Number and Numerical Operations
- II Geometry and Measurement
- III Patterns and Algebra
- IV Data Analysis, Probability, and Discrete Mathematics

Within these standards there are a number of specific clusters. Appendix A identifies all of those. Based on the Spring 2002 operational test, test points come approximately equally from the four content areas such that there are 12 points for each. (This differs from the February 1998 document which shows a 15%, 25%, 30%, and 30% respective weighting.)

The test for Mathematics consists of 48 points:

1 point for each of 30 multiple-choice items = 30

3 points for each of 6 open-ended items (two readers' scores averaged) = 18

Tests are assembled using p -values and means of items from field-testing and are equated using a Rasch model with theta for the appropriate cut scores. Table 2.1.1 summarizes the total points possible for each of the content areas of the operational NJ HSPA administered March 2015.

Table 2.1.1: 2015 March NJ HSPA Total Points By Content Area

Language Arts Literacy		Points	
Total		54	
Writing		18	
Picture Prompt		6	
Persuasive Prompt		12	
Reading		36	
Working with Text		10	
Analyzing Text		26	
Mathematics		Points	Problem Solving Points
Total		48	42
I Number and Numerical Operations		7	5
II Geometry and Measurement		12	12
III Patterns and Algebra		15	11
IV Data Analysis, Probability, and Discrete Mathematics		14	14

2.2 Development of Test Items

The process for developing an item requires a two-year cycle in which the items are

- Written to HSPA standards;
- Reviewed by MI content experts;
- Reviewed by state content experts;
- Reviewed by New Jersey teachers and a sensitivity committee before allowing the item to be field-tested;
- Field-tested;
- Reviewed before scoring by a range-finding committee involving state educators; and
- Reviewed again by state content experts, New Jersey teachers, and a sensitivity committee after field-testing.

Only an item that has been found acceptable at every stage of the two-year cycle is entered into the bank for possible use on an operational test. Although statistical data on test items play an essential role, this cycle of development employs a due process model of validity. This model relies on the expertise of educators participating in the test development process. The strength of this process is dependent on the degree to which the following critical components are integrated into the test development process:

- Recruitment of expert educators familiar with the state’s content specifications and population for the assessment;
- Training of item writers and expert reviewers on item writing specifications, content specifications, and the goals and functions of the assessment;

- Careful consideration of individual items by experts to assess the degree to which the items measure the knowledge, skills, and abilities the assessment is intended to measure, along with opportunities to reject or revise items per committee recommendation; and
- Careful consideration of sensitivity issues by experts to guarantee that performance on the item is related to classroom achievement and not cultural or social experiences outside the classroom, along with opportunities to reject or revise items per committee recommendation.

At Measurement Incorporated™, item writers, under the supervision of content experts, are instructed on the state specifications and item types necessary for the tests. They are trained on the HSPA content specifications and directed to write original items tailored to NJ content standards. Content expert reviewers at MI validate (or not) the initial form and coding of items written to meet HSPA content standards. At this point in the process, some items are rejected from further consideration on the grounds that the items are not tied closely enough to HSPA standards or are not at an appropriate level of difficulty.

In reviewing items, NJ educators review information beyond the item wording and scoring rubric. In MATH, teachers validate the assignment of each item to a NJ content specification Standard and Cluster, using the same standards used by the item writers. Teachers also review an item assignment to a Knowledge or Problem-solving category. LAL committee members review the type of passage and skill cell of each LAL item. For all content review meetings, MI furnished reviewers with copies of the NJ skill code (LAL) and Cluster-Standard (MATH) sheets to allow committee members to validate assignment of items to NJ content standards. Reviewers may accept or revise an item coding assignment, or reject an item as not fulfilling any specific part of the content specifications. For each item, both committees also rate each item for a level of difficulty.

Although the content committees are trained to recognize possible bias or lack of cultural sensitivity in test items, a separate sensitivity committee meets to review Language Arts Literacy passages before field-testing to identify potential item bias. After field-testing, the same committee reviews all multiple-choice items flagged as Mantel-Haenszel “C” items (probable DIF) in Language Arts Literacy and Mathematics and all open-ended items using student data disaggregated by demographic group for both tests. Like the content committee, the sensitivity committee has the power to reject an item. If either the sensitivity committee or content committee rejects an item, it is considered rejected. If one requires that the item be revised, that decision outweighs an acceptance by the other committee.

Each field test multiple-choice item has a Mantel-Haenszel statistic for each of three comparisons that New Jersey student population will support. A White/African American, White/Hispanic, and Male/Female comparison for each multiple-choice item is done with adequate sample sizes for the focus groups (e.g. African Americans, Hispanic, or Females). It is extremely rare that items with Mantel-Haenszel “C” categorizations are used for the operational assessments; it only occurs when there are no other items available to cover the NJ HSPA content standards.

At item review session, items are presented one-per-page with the footer below.
 Footer (Common to both content areas)

Sensitivity			Content		
*Comments:			*Comments:		
Sensitivity Issue	Yes	No	Meets Specifications	Yes	No
If yes, identify category and explain*			Appropriate Difficulty	Yes	No
			Accurate Coding	Yes	No
Definitely Use			Definitely Use		
Revise and Use With Approval			Revise and Use With Approval		
Revise and Resubmit			Revise and Resubmit		
Do Not Use*			Do Not Use*		

At the bottom of each footer is a place for committee members to sign off on their decision:

Sensitivity Sign-off	Date	Content Chairperson's Signature	Date
----------------------	------	---------------------------------	------

This is a critical step in the item review process as it records, item by item, the appropriateness of each item for the assessment. Only an item approved by both committees can be field-tested.

The Field Test Form Distribution

Before Spring of 2002, the NJ DOE developed items for the HSPA using a standalone field-test format. Beginning with the first operational administration in Spring of 2002, the NJ DOE began embedding field test items for Mathematics and Language Arts Literacy. This year no items were field tested for mathematics while all forms for LAL were used for equating purposes for next year. Due to the lack of the need for field testing new items the NJ HSPA Spring 2015 assessments had only one form.

2.3 Item Review Process

Each summer following a field test, the NJ DOE conducts a statistical analysis review session with New Jersey teachers. The teachers on the content and sensitivity committees review the items and evaluate the performance of the items based on field test data. No item review was conducted in 2015; however, the process was used in previous year's concerning all the items that were included on the NJ HSPA 2015 operational assessments. The following variables were included:

- Item ID
- N-count
- *p*-value
- Biserial
- % answering each option (A-D) and omits
- *p*-value for bottom 20%
- *p*-value for top 20%
- % of Whites answering each option (A-D) and omits; N-count for Whites
- % of Blacks answering each option (A-D) and omits; N-count for Blacks
- % of Hispanics answering each option (A-D) and omits; N-count for Hispanics
- % of Males answering each option (A-D) and omits; N-count for Males
- % of Females answering each option (A-D) and omits; N-count for Females
- Total Reading Score for students taking that form
- Total Writing Score for students taking that form
- OE items' mean score
- Correlation of each OE item with total reading score
- Correlation of each OE item with total writing score
- OE item score distribution, frequency, percent, mean, and standard deviation for total group
- OE item score distribution, frequency, percent, mean, and standard deviation for Whites
- OE item score distribution, frequency, percent, mean, and standard deviation for Blacks
- OE item score distribution, frequency, percent, mean, and standard deviation for Hispanics
- OE item score distribution, frequency, percent, mean, and standard deviation for Males
- OE item score distribution, frequency, percent, mean, and standard deviation for Females
- OE Base Test mean for above students (Reading and Writing combined)
- OE Base Test mean for state (Reading and Writing Combined)
- OE Reason Code Distribution for Nonscorable responses (Number of 6, 7, 8, and 9s)
- Mantel-Haenszel statistics

For the meeting, teachers are provided with a training session on how to interpret these statistics. To draw their attention to items that may be problematic, several flags are used. The flags include:

- Difficulty flag to indicate that an item has a *p*-value less than .25 or greater than .95
- Correlation flag to indicate an item that has an item-total correlation of less than .25
- Mantel-Haenszel flags to indicate any group comparison flagged as "C" using the standard

ETS coding of Mantel-Haenszel results into A, B, C.

2.4 Item Use

At the meetings, teachers were presented with forms similar to those used at initial item development meetings:

ITEM CODE AND KEY		Admin: March 2009	Form:	Position:
Sensitivity		Content		
*Comments		*Comments		
Sensitivity Issue	<input type="checkbox"/> Yes <input type="checkbox"/> No	Appropriate Difficulty	<input type="checkbox"/> Yes <input type="checkbox"/> No	
If a sensitivity issue, explain*		P-value = Biserial =		
Mantel-Haenszel Category C <input type="checkbox"/> W-AA <input type="checkbox"/> W-H <input type="checkbox"/> M-F				
<input type="checkbox"/>	Definitely Use			<input type="checkbox"/>
<input type="checkbox"/>	Revise and Use With Approval **			<input type="checkbox"/>
<input type="checkbox"/>	Revise and Re-Field Test			<input type="checkbox"/>
<input type="checkbox"/>	Do Not Use *			<input type="checkbox"/>

Sensitivity Sign-Off	Date	Content Chairperson's Signature	Date
----------------------	------	---------------------------------	------

** Requires director's approval

- **Accept (Definitely Use):** All content related issues (importance, thematic, grammar, clarity, accuracy, validity, sound measurement, grade-appropriate), all statistical criteria, and all sensitivity issues have been met or exceeded and the item appears suitable for operational use.
- **Revise (Revise and Re-Field Test):** One or more of the content related issues have not been met or the item needs minor changes to make it acceptable. Reviewers provide recommendations on changes to be made to the item that will make the item suitable for re-field testing.
- **Reject (Do Not Use):** Several content related issues, statistical criteria, or sensitivity issues have not been met, or are suspect, or need radical changes to make the item acceptable. In such cases, the item may be vague or ambiguous, inappropriate, or not clearly related to the text or to the standard. Without severe modifications, it is unlikely to be salvaged. Reviewers provide comments as to why the item should be rejected.
- **Revise and Use With Approval:** A very minor content related issue needs to be resolved and the NJ DOE content representative feels it is minor enough to use operationally without re-field testing.

Only items designated as **revise and use with approval** or **accepted** by both committees are added to the item bank for possible use on future operational tests.

2.5 Quality Control for Test Construction

Measurement Incorporated and the NJ DOE work very closely together to ensure that the content and editorial quality of the test booklets meet or exceed the state expectations. This requires consistent vigilance and quality control checks during the test booklet assembly process.

The operational test is assembled from the approved HSPA test design using field-tested items that are proven valid and fair to all students. Test booklets are assembled using approved general and HSPA style guidelines. The typeset test booklets are then proofread by two editorial staff members for typographical and format errors and also to determine if the version of the item used is consistent with the version field-tested. Upon completion of this internal review, the test booklet is sent to NJ DOE for a typeset review. MI then responds to any revisions required by NJ DOE, and this process is repeated until NJ DOE approves the test booklet. Once NJ DOE approves the operational test booklet, it is then sent to a proofreader external to MI for an independent review. Any recommendations by the external reviewer are given to NJ DOE for consideration. Any desired changes are then made. The final approved version of the test booklet is then converted to a pdf file for printing. The pdf version of the test booklet is proofread by editorial staff before submitting to the printing manager. The first copies of the production run of the test booklet are reviewed for possible problems by the MI project management staff and the editorial staff.

Ancillary test materials are subject to the same consistent vigilance and quality control. All typeset copies are first proofread by at least two editorial staff members before submitting them to NJ DOE for a typeset review. Materials must be approved by NJ DOE before printing. Approved versions of the ancillary materials are then converted to pdf files for printing.

All accommodated materials are also subject to consistent vigilance and quality control at all stages. The large print test and supporting materials are subject to the same assembly quality control discussed previously. The Braille translation of the test and supporting materials is performed by an independent, certified translation agency. The large print and Braille versions of the test materials are then submitted to NJ DOE for review by specialists from the state Commission for the Blind. Revisions to the materials are made based on recommendations from these state specialists, and then the accommodated materials are sent to production.

Part 3: Test Administration

3.1 Participation

The 2015 March administration was for students who were retesting after previously scoring below the state's graduation testing requirement of 200 or above. Prior to 2015 the October and March administrations were for the following students who have not yet passed all required sections of the HSPA:

- Retained eleventh-grade students;
- Twelfth-grade students;
- Retained twelfth-grade students;
- Adult high school students; and
- Returning students.

3.2 Test Security Procedures

The following materials are considered secure during the test administration.
(All materials are shrink-wrapped in packages of ten.)

- Test Booklets: **Must** remain secure at **all** times.
- Answer Folders: **Used** answer folders **must** be kept secure at all times.
- Persuasive Writing Task Folders (Purple): The packages **must not** be opened until the Persuasive Writing Task Folders are distributed to students for the regular Language Arts Literacy – Day 2 administration.
- Make-Up Persuasive Writing Task Folders (Blue): The Make-Up Writing Task Folders are contained in the District Overage Box. They are provided for students who are absent from the regular administration of Language Arts Literacy – Day 2. Do **not** distribute the Make-Up Persuasive Writing Task Folders to school test coordinators until the make-up week.

The HSPA test booklet and its contents are secure materials. It *CANNOT* be read, reviewed or discussed by staff, either orally or in writing, or copied either wholly or in part, for any purpose. It is the responsibility of the district to guarantee the security of the test materials. Security breaches may have financial consequences for the district, professional consequences for staff, and disciplinary consequences for students.

All security breaches involving missing test booklets will be investigated by the New Jersey Department of Education.

The items and passages contained in the HSPA test booklet must remain confidential because some of the test items will reappear in future versions of the test. This is done to maintain the

stability of the test item pool over time from a technical perspective, and to enable comparisons to be made from one year to the next.

- All district and school personnel must be informed of the HSPA test security procedures prior to the test administration, including those personnel not directly involved in administering the test.
- All examiners must sign a New Jersey Department of Education Statewide Assessments Test Security Agreement including staff trained to be substitute examiners. The agreement **must** be signed immediately after the mandatory training session conducted by the school or district test coordinator. Signed agreements **must** be kept in the district for one year after completion of a test administration cycle. Do **not** return the forms to the test contractor.
- All test materials **must** be stored in a securely locked location **at all times**. This secure area should be accessible only to individuals who are authorized by the school test coordinator.
- Under no circumstances are students permitted to have cell phones, MP3 players, or other unauthorized electronics in their possession during testing. **Students found to have unauthorized electronics in a testing room will receive a V2 (void) for that test section.**
- The plastic wrapping around the test booklets may be opened the day before testing to prepare test materials for distribution to examiners the next day.
- Do **not** break the plastic wrapping around the Persuasive Writing Task Folders until Day 2 of the Language Arts Literacy administration.
- Make-up Persuasive Writing Task Folders will be included with the district overage materials. It is the responsibility of the district test coordinator to distribute the Make-Up Persuasive Writing Task Folders to school test coordinators the week of the make-up administration.
- Examiners and proctors may **not** coach or assist students other than reading the directions outlined in the *Examiner Manual*.
- Examiners, proctors, and other school personnel are **not** permitted to discuss or disclose any test items, reading passages, or writing prompts before, during, or after the test administration.
- The chief school administrator or designee must sign for the initial shipment of test materials and present the HSPA Authorization to Receive Secure Test Materials Form, signed by the chief school administrator or designee, to the NCD Package Express agent when the materials are delivered. The authorization form is not necessary to receive supplementary shipments which will be shipped via FedEx directly to the district test coordinator.
- Each test booklet and answer folder has a unique identification number. Students **must** use the same test booklet and answer folder for the entire test. Students **must** print their names on the front cover of their test booklets and record their test booklet numbers on their answer

folders.

- The following forms are provided and **must** be used while test materials are in the district:

Authorization to Receive Secure Test Materials Form

District Receipt Form

School Security Checklists – Regular and Make-up Administration

School Security Checklists – District Overage

School Security Checklists – Supplementary Shipment (if applicable)

District Return Form

Irregularity Report

- The principal and the chief school administrator or official designees must review and sign the completed Header Sheets. These signatures affirm that the answer folder return totals are correct and that all HSPA test administration procedures outlined in the manuals have been followed.
- Test booklets should only be handled by authorized staff that include: the district test coordinator, school test coordinators, examiners, and the personnel who are responsible for inventorying, distributing, collecting, and returning test booklets. Proctors are **not** permitted to handle test booklets.
- Strict test security must be maintained. The Office of Assessments, in cooperation with county offices of education, will monitor all testing and security procedures.

3.3 Test Administration Procedures

The HSPA is administered on three consecutive days (Tuesday, Wednesday, and Thursday.) The testing should not be scheduled after an athletic event or an assembly. All test schedules should be checked with the appropriate school officials to ensure that other school activities do not interfere with the test administration.

- **All testing MUST be scheduled in the morning except for adult high school, homebound, and bedside students, and for students attending out-of-district placements who are tested at that placement by staff from the student's home district.**
- The district and school test coordinators are responsible for scheduling times and places for regular and make-up testing and for ensuring that all testing is completed according to the procedures and schedule described in the *District/School Test Coordinator* manual and in the *Examiner Manual*.
- Students who are required to test but are absent for the regular test administration must be tested on the make-up dates.
- Students whose answer folders are voided during testing are considered to have attempted the

test section. They may NOT retake or resume taking the voided test section during the make-up. They must wait until the next HSPA administration: October for answer folders voided in March; and March for answer folders voided in October.

- Students who begin a section of the test and do not complete it during the specified testing time **cannot** complete the test section during the make-up period or any other time **unless** additional time is specified in a student’s IEP or 504 plan.

STUDENT ROSTERS

Under the direction of the district test coordinator, the school coordinator must prepare a student roster for each examiner. Each roster should list the names of the students who each examiner will supervise during testing, the names of proctors assisting the examiner, and the room number. Ensure that each roster contains any information that must be gridded on a student’s answer folder since some students may not have a HSPA label or may have information missing from their HSPA label. In addition, **all** test modifications and accommodations for SE and 504 students must be indicated as applicable.

Refer to the “School Use Only” section of the *Examiner Manual* for detailed information regarding the gridding of the answer folder. Distribute the student rosters to examiners in advance of testing to allow the examiners to prepare for the test administration. Examiners should also use the rosters to take student attendance for the purpose of planning for make-up testing. Examiners must return the student rosters to the school coordinator immediately after testing has been completed. The school test coordinator should keep a copy of the rosters and return the originals to the district test coordinator. Use these rosters to identify answer folders to be kept for the make-up week.

Advance Announcements

It is important that teachers have advance notice of the testing schedule so they may adjust lesson plans and personal schedules.

Instruct examiners to advise their students about the time and location of the test administrations and a brief explanation about the test. Students **must** bring two #2 pencils if the school does not provide them. No pens of any kind are allowed. All sections of the answer folder **must** be marked in pencil.

Students **must** be notified that they may **not** use dictionaries or other reference materials during testing, and that they are **not** permitted to have cell phones, MP3 players, or other unauthorized electronics in any testing room, and that **any student found to have unauthorized electronics in his or her possession while in a testing room will have his or her answer folder voided (V2) for that test section.**

OCTOBER TEST ADMINISTRATION TIMES

- The approximate administration times for the October HSPA are as follows:

Mathematics	2 hours, 32 minutes
Language Arts Literacy – Day 1	1 hour, 36 minutes
Language Arts Literacy – Day 2	2 hours, 1 minute

This schedule includes time for gridding student information on the answer folder the first day of testing, distributing the test materials, reading directions, test taking, providing breaks for the students, and collecting materials after testing.

- The number of test questions in each section is as follows:

Mathematics	30 multiple-choice and 6 open-ended questions
Language Arts Literacy	20 multiple-choice and 4 open-ended questions and 2 writing tasks

MARCH TEST ADMINISTRATION TIMES

- The approximate administration times for the March HSPA are as follows:

Mathematics	2 hours, 32 minutes
Language Arts Literacy – Day 1	2 hours, 10 minutes to 2 hours, 40 minutes
Language Arts Literacy – Day 2	2 hours, 1 minute

This schedule includes time for gridding student information on the answer folder the first day of testing, distributing the test materials, reading directions, test taking, providing breaks for the students, and collecting materials after testing.

- The number of test questions in each section is as follows:

Mathematics	30 multiple-choice and 6 open-ended questions
Language Arts Literacy	20 multiple-choice and 4 open-ended questions, 2 writing tasks, and field-test questions and writing tasks

CALCULATOR REQUIREMENTS

- Students are required to have calculators during the Mathematics Section.
- Instructional manuals or function reference sheets must be removed prior to testing and calculator memories must be cleared both before and after the Mathematics administration.
- A student may choose to use either a personal calculator or one that is provided by the

district. Students should use the calculators they use on a regular basis either in the classroom or at home.

- The calculators must have algebraic logic, the ability to do powers and roots of any degree, and at least one memory cell. Computers (laptops, palmtops, etc.), pocket organizers, and calculators with QWERTY (i.e., typewriter) keyboards, or built-in computer algebra systems (CAS) are **not** acceptable.

3.4 Test Accommodations

Accommodations and Modifications of Test Administration Procedures for Special Education Students and Students Eligible under Section 504 of the Rehabilitation Act of 1973

In accordance with the Individuals with Disabilities Education Act (IDEA) of 1997 and the No Child Left Behind Act, all students with disabilities must participate in state assessments. Students with disabilities who are in the eleventh grade must participate in the High School Proficiency Assessment (HSPA) or the Alternate Proficiency Assessment (APA). The Individual Educational Plan (IEP) team for each student determines which assessment a student will take in accordance with the NJAC 6A:14. The regulations state that each student with disabilities must take the general state assessment, which is the HSPA for eleventh-graders, unless the student has not been instructed in any of the knowledge and skills tested and cannot complete any of the types of tasks on the HSPA. The IEP team must determine which assessment the student will take for each content area assessed. Students who do not take the HSPA in one or more content areas assessed must participate in the APA in that/those content area(s).

The IEP team must also determine if the student who is taking the HSPA in one or both content areas will be required to pass the HSPA in those content areas in order to graduate. **If a student's IEP states that he or she must pass the HSPA in one or more content areas in order to graduate, and the student does not pass the specified content areas in eleventh grade, the student will participate in the Alternative High School Assessment (AHSA) during the twelfth grade. Students with disabilities participate in the AHSA only if they are required to pass the HSPA in one or more content areas in order to graduate.** The IEP team may also determine that the student does not have to take the HSPA a second or third time while he or she is participating in the AHSA, if taking the HSPA again would be detrimental to the student.

Districts may use accommodations and modifications of test administration procedures when administering the HSPA to special education students or to students eligible under Section 504 of the Rehabilitation Act of 1973. Accommodations and modifications in the areas listed below may be used separately or in combination.

Decisions about participation and accommodations and modifications are made by the IEP or 504 team. Information about test content and item types from the test specifications booklets can be used to make this determination.

Accommodations and modifications of test administration procedures for students eligible under IDEA or Section 504 must be specified in the student's IEP or 504 plan. Accommodations and modifications must be consistent with the instruction and assessment procedures used in the student's classroom. Students eligible for accommodations and modifications under Section 504 may not be classified but do have a permanent or temporary impairment in a major life function (for example: performing manual tasks, walking, seeing, hearing, speaking, etc.).

Advanced planning is integral to implementing accommodations and modifications effectively and ensuring that the security of test materials is maintained.

Accommodations and modifications must be recorded on the student's answer folder using unique codes.

Acceptable Accommodations and Modifications

Code

A. Setting Accommodations

1. Administering the assessment:
 - a. individually in a separate room
 - b. in a small group in a separate room
 - c. in a resource room
 - d. in a special education classroom
 - e. at home or in a hospital (this will depend on the nature of the assessment task)
2. Seating the student in the front of the room near the examiner or proctor
3. Seating the student facing the examiner or proctor
4. Providing special lighting
5. Providing special furniture e.g., desks, trays, carrels

B. Scheduling Accommodations

1. Adding time as needed
2. Providing frequent breaks
3. Terminating a section of the test when a student has indicated that he/she has completed all the items he/she can. The examiner must ensure that the student has attempted all items in a section since items are not ordered by difficulty. When this accommodation is used, the test must be administered in a small group or individually to avoid distraction.

C. Test Materials Modifications

1. Administering the large-print version of the test
2. Administering the Braille version of the test

D. Test Procedures Modifications

1. Administration modifications
 - a. reading directions aloud
 - b. reading test items aloud (do NOT read aloud or sign the reading passages in Language Arts Literacy– the reading items may be read or signed); ONLY the teacher who must read aloud or sign the test items is permitted to have a test booklet assigned to him/her for this task

- c. providing and ensuring that amplification (hearing aid and/or FM system) is in working order
 - d. using a sign language or cued speech interpreter to sign or cue the directions or test items but NOT the reading passages
 - e. masking a portion of the test booklet and/or answer folder to eliminate visual distractors or providing reading windows
 - f. repeating, clarifying, or rewording directions ONLY
 - g. providing written directions on a separate sheet or transparency
 - h. using an examiner who is familiar with the student
 - i. using an examiner who can communicate fluently in sign language (American Sign Language or a form of Manually Coded English)
 - j. providing manipulatives for math items e.g. number line, counting chips, abacus (for NJASK 3–8 ONLY)
 - k. using graph paper for HSPA Mathematics (all students are permitted graph paper for NJASK 3–8)
 - l. using a Braille ruler and talking calculator or large-face calculator
 - m. using tactile or visual cues for deaf or hard of hearing students to indicate time to begin, time remaining, and time to end a particular part of the test
 - n. using calculators for NJASK 3–8 Mathematics (all students are permitted calculators for HSPA)
2. Response modifications
- a. having an examiner record the student’s identification information on the test booklet and/or answer folder
 - b. dictating oral responses to a scribe (examiner or proctor who writes from dictation)
 - c. using a Braille writer to record responses
 - d. signing responses to a sign language interpreter (student must indicate all punctuation and must spell all key words)
 - e. recording responses on a word processor (all editorial functions MUST be disabled)
 - f. providing an augmentative communication device
 - g. using a larger diameter or modified special grip #2 pencil
 - h. circling answers in the test booklet (the examiner subsequently transfers the answers to the answer folder; for NJASK 3–4, the examiner bubbles the student’s answer choice in the scanable test booklet)
 - i. allowing separate additional continuation pages for writing tasks

Other Considerations

Ensure that:

- a. any medication has been appropriately adjusted so it will not interfere with the student’s functioning.
- b. eyeglasses are used, if needed.
- c. hearing aids, FM systems, augmentative communication devices, word processors, or other types of equipment are functioning properly.
- d. source and strength of light are appropriate.
- e. all students can clearly see and hear the examiner.
- f. all deaf or hard of hearing students who communicate aurally/orally are watching

the examiner when instructions are given.

For students responding on separate sheets of paper:

- a. Responses to open-ended items and writing tasks, which are written or typed on separate sheets of paper by students eligible for this accommodation, must be placed in the **YELLOW SE/504 Accommodations & Braille/Large-Print Envelope**.
 1. The answer folder and separate sheets must be clipped together and placed beneath a separate grade-level Header Sheet.
 2. Write on the top of each separate sheet the student's: name, answer folder number, birth date, district name and code, and school name and code. **If these procedures are not followed, these responses cannot be linked to the student's responses in the answer folder and the student will receive incomplete scores.**
- b. Copies of these sheets must be made and retained on file by the school district until scores are received and verified.

For large-print test administration:

- a. Students taking the large-print test:
 1. Will mark their answers in the large-print test booklet and the examiner will transcribe the responses onto the regular answer folder provided.
 2. May be instructed to skip some items identified in the large-print supplemental instructions. The spaces for these items must be left blank on the student's answer folder included in the large-print kit.
- b. Answer folders used for the large-print test:
 1. Must be sent to Measurement Incorporated in the **YELLOW SE/504 Accommodations & Braille/Large-Print Envelope**.
 2. Grid SE/504 ACCOMM = A, B, C, and D. A separate grade-level Header Sheet must be completed for answer folders containing the multiple-choice responses.
- c. For dictations and responses recorded on separate sheets of paper:
 1. Students who dictate responses to open-ended items and writing tasks must indicate all punctuation and must spell all key words.
 2. Responses to open-ended items and writing tasks recorded on separate sheets of paper must be clipped to the answer folder. The student's name, answer folder number, and birth date as well as the district and school names and codes must be recorded on each of the separate sheets attached.

3. The district test coordinator must retain a duplicate of all open-ended item responses and writing task responses until scores are reported to and reviewed by district staff.

For Braille test administration:

- a. Students taking the Braille test:
 1. Will dictate their answers to the examiner and/or use a device that produces Braille.
 2. May be instructed to skip some items identified in the Braille supplemental instructions. The spaces for these items must be left blank on the student's answer folder included in the Braille kit.
 3. Should be instructed to bring a Braille ruler and a talking calculator to the test session.
- b. Answer folders used for the Braille test:
 1. Must be sent to Measurement Incorporated in the **YELLOW SE/504 Accommodations & Braille/Large-Print Envelope**.
 2. Grid SE/504 ACCOMM = A, B, C, and D. A separate grade-level Header Sheet must be completed for answer folders containing the multiple-choice responses.
- c. For dictations and responses recorded in Braille:
 1. Students who dictate responses for the open-ended items and writing tasks must indicate all punctuation and must spell all key words.
 2. Responses to the writing task and open-ended questions recorded in Braille must be transcribed. The transcriptions of the Brailled responses, along with the student's own Brailled work, must be clipped to the answer folder. The student's name, answer folder number, and birth date as well as the district and school names and codes must be recorded on each of the separate sheets attached.
 3. The district test coordinator must retain a duplicate of all open-ended item responses and writing task responses until scores are reported to and reviewed by district staff.

For students who communicate using sign language:

- a. An interpreter will be needed to interpret oral directions and test items (do NOT interpret the reading passages in Language Arts Literacy). The interpreter must be able to communicate in the mode used by the student, American Sign Language or a form of Manually Coded English, depending upon the student's communication system. The interpreter must be instructed to interpret so as not to give the answer to the student through the use of a particular sign or finger spelling.
- b. Students using American Sign Language for open-ended item responses and writing task responses will sign the responses to the interpreter who will interpret them into spoken English, and a scribe will record the responses in the answer folder.
- c. Students using Signed English or cued speech will sign or cue the responses to the interpreter who will transliterate (word for word) into spoken English, and a scribe will record the responses in the answer folder.

For students needing a scribe:

The student will dictate the response at first without indicating punctuation and spelling. This allows the student to get his/her ideas on paper. The scribe must write what is said, without capitals or punctuation and without the student seeing, but with correct spelling. Once the student has finished the response, the scribe must identify key words and ask the student to spell them. The scribe can underline the words and write the student's spelling above the word or write the student's spelling at the bottom of the page. The scribe must then show the piece to the student and ask him/her to indicate what words should be capitalized and where punctuation should occur. The student also has the opportunity to edit the piece at this point. The scribe must not advise or lead the student in any way.

Table 3.4.1: Special Education (Setting Accommodation)

	Language Arts Literacy							Mathematics						
	Number of Valid Scale Scores	Partially Proficient		Proficient		Advanced Proficient		Number of Valid Scale Scores	Partially Proficient		Proficient		Advanced Proficient	
		No.	%	No.	%	No.	%		No.	%	No.	%	No.	%
Special Education	486	371	0.76	112	0.23	3	0.01	1,038	902	0.87	135	0.13	1	0.00
Setting Accommodation	364	270	0.74	92	0.25	2	0.01	824	710	0.86	113	0.14	1	0.00
Auditorily Impaired	0	0	.	0	.	0	.	3	3	1.00	0	0.00	0	0.00
Autistic	9	6	0.67	3	0.33	0	0.00	12	11	0.92	1	0.08	0	0.00
Cognitively Impaired – Mild	12	11	0.92	1	0.08	0	0.00	18	18	1.00	0	0.00	0	0.00
Cognitively Impaired – Moderate	1	1	1.00	0	0.00	0	0.00	2	1	0.50	1	0.50	0	0.00
Cognitively Impaired – Severe	1	1	1.00	0	0.00	0	0.00	1	1	1.00	0	0.00	0	0.00
Communication Impaired	23	20	0.87	3	0.13	0	0.00	50	45	0.90	5	0.10	0	0.00
Emotionally Disturbed	41	25	0.61	14	0.34	2	0.05	63	52	0.83	10	0.16	1	0.02
Multiply Disabled	40	34	0.85	6	0.15	0	0.00	61	53	0.87	8	0.13	0	0.00
Deaf-Blindness	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Orthopedically Impaired	2	2	1.00	0	0.00	0	0.00	4	4	1.00	0	0.00	0	0.00
Other Health Impaired	51	33	0.65	18	0.35	0	0.00	144	118	0.82	26	0.18	0	0.00
Social Maladjustment	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Specific Learning Disability	173	129	0.75	44	0.25	0	0.00	451	389	0.86	62	0.14	0	0.00
Traumatic Brain Injury	1	0	0.00	1	1.00	0	0.00	5	5	1.00	0	0.00	0	0.00
Visually Impaired	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Eligible for Speech-Language Services	1	1	1.00	0	0.00	0	0.00	2	2	1.00	0	0.00	0	0.00
Multiple Grids and Default	9	7	0.78	2	0.22	0	0.00	8	8	1.00	0	0.00	0	0.00
Section 504	26	11	0.42	13	0.50	2	0.08	160	123	0.77	36	0.23	1	0.01

Table 3.4.2: Special Education (Scheduling Accommodations)

	Language Arts Literacy							Mathematics						
	Number of Valid Scale Scores	Partially Proficient		Proficient		Advanced Proficient		Number of Valid Scale Scores	Partially Proficient		Proficient		Advanced Proficient	
		No.	%	No.	%	No.	%		No.	%	No.	%	No.	%
Special Education	486	371	0.76	112	0.23	3	0.01	1,038	902	0.87	135	0.13	1	0.00
Scheduling Accommodations	358	265	0.74	91	0.25	2	0.01	834	718	0.86	116	0.14	0	0.00
Auditorily Impaired	0	0	.	0	.	0	.	3	3	1.00	0	0.00	0	0.00
Autistic	9	6	0.67	3	0.33	0	0.00	11	10	0.91	1	0.09	0	0.00
Cognitively Impaired – Mild	12	11	0.92	1	0.08	0	0.00	18	18	1.00	0	0.00	0	0.00
Cognitively Impaired – Moderate	1	1	1.00	0	0.00	0	0.00	2	1	0.50	1	0.50	0	0.00
Cognitively Impaired – Severe	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Communication Impaired	22	19	0.86	3	0.14	0	0.00	52	46	0.88	6	0.12	0	0.00
Emotionally Disturbed	40	25	0.63	13	0.33	2	0.05	62	52	0.84	10	0.16	0	0.00
Multiply Disabled	39	33	0.85	6	0.15	0	0.00	58	53	0.91	5	0.09	0	0.00
Deaf-Blindness	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Orthopedically Impaired	2	2	1.00	0	0.00	0	0.00	4	4	1.00	0	0.00	0	0.00
Other Health Impaired	53	35	0.66	18	0.34	0	0.00	148	121	0.82	27	0.18	0	0.00
Social Maladjustment	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Specific Learning Disability	176	132	0.75	44	0.25	0	0.00	468	402	0.86	66	0.14	0	0.00
Traumatic Brain Injury	1	0	0.00	1	1.00	0	0.00	5	5	1.00	0	0.00	0	0.00
Visually Impaired	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Eligible for Speech-Language Services	1	0	0.00	1	1.00	0	0.00	1	1	1.00	0	0.00	0	0.00
Multiple Grids and Default	2	1	0.50	1	0.50	0	0.00	2	2	1.00	0	0.00	0	0.00
Section 504	26	11	0.42	13	0.50	2	0.08	169	130	0.77	38	0.22	1	0.01

Table 3.4.3: Special Education (Test Materials Modifications)

	Language Arts Literacy							Mathematics						
	Number of Valid Scale Scores	Partially Proficient		Proficient		Advanced Proficient		Number of Valid Scale Scores	Partially Proficient		Proficient		Advanced Proficient	
		No.	%	No.	%	No.	%		No.	%	No.	%	No.	%
Special Education	486	371	0.76	112	0.23	3	0.01	1,038	902	0.87	135	0.13	1	0.00
Test Materials Modification	21	19	0.90	2	0.10	0	0.00	21	18	0.86	3	0.14	0	0.00
Auditorily Impaired	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Autistic	1	0	0.00	1	1.00	0	0.00	0	0	.	0	.	0	.
Cognitively Impaired – Mild	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Cognitively Impaired – Moderate	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Communication Impaired	1	1	1.00	0	0.00	0	0.00	1	1	1.00	0	0.00	0	0.00
Emotionally Disturbed	2	2	1.00	0	0.00	0	0.00	1	1	1.00	0	0.00	0	0.00
Multiply Disabled	2	2	1.00	0	0.00	0	0.00	2	2	1.00	0	0.00	0	0.00
Deaf-Blindness	3	3	1.00	0	0.00	0	0.00	3	2	0.67	1	0.33	0	0.00
Orthopedically Impaired	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Other Health Impaired	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Specific Learning Disability	3	3	1.00	0	0.00	0	0.00	4	3	0.75	1	0.25	0	0.00
Traumatic Brain Injury	2	2	1.00	0	0.00	0	0.00	4	3	0.75	1	0.25	0	0.00
Visually Impaired	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Eligible for Speech-Language Services	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Multiple Grids and Default	7	6	0.86	1	0.14	0	0.00	6	6	1.00	0	0.00	0	0.00
Section 504	0	0	.	0	.	0	.	1	1	1.00	0	0.00	0	0.00

Table 3.4.4: Special Education (Test Procedure Modifications)

Special Education	Language Arts Literacy							Mathematics						
	Number of Valid Scale Scores	Partially Proficient		Proficient		Advanced Proficient		Number of Valid Scale Scores	Partially Proficient		Proficient		Advanced Proficient	
		No.	%	No.	%	No.	%		No.	%	No.	%	No.	%
Special Education	486	371	0.76	112	0.23	3	0.01	1,038	902	0.87	135	0.13	1	0.00
Test Procedure Modifications	319	239	0.75	79	0.25	1	0.00	704	607	0.86	97	0.14	0	0.00
Auditorily Impaired	0	0	.	0	.	0	.	2	2	1.00	0	0.00	0	0.00
Autistic	8	5	0.63	3	0.38	0	0.00	11	10	0.91	1	0.09	0	0.00
Cognitively Impaired – Mild	11	10	0.91	1	0.09	0	0.00	17	17	1.00	0	0.00	0	0.00
Cognitively Impaired – Moderate	1	1	1.00	0	0.00	0	0.00	2	1	0.50	1	0.50	0	0.00
Cognitively Impaired – Severe	1	1	1.00	0	0.00	0	0.00	1	1	1.00	0	0.00	0	0.00
Communication Impaired	19	17	0.89	2	0.11	0	0.00	42	37	0.88	5	0.12	0	0.00
Emotionally Disturbed	36	24	0.67	11	0.31	1	0.03	56	48	0.86	8	0.14	0	0.00
Multiply Disabled	38	32	0.84	6	0.16	0	0.00	58	52	0.90	6	0.10	0	0.00
Deaf-Blindness	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Orthopedically Impaired	2	2	1.00	0	0.00	0	0.00	3	3	1.00	0	0.00	0	0.00
Other Health Impaired	46	31	0.67	15	0.33	0	0.00	124	100	0.81	24	0.19	0	0.00
Social Maladjustment	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Specific Learning Disability	147	109	0.74	38	0.26	0	0.00	375	323	0.86	52	0.14	0	0.00
Traumatic Brain Injury	1	0	0.00	1	1.00	0	0.00	4	4	1.00	0	0.00	0	0.00
Visually Impaired	0	0	.	0	.	0	.	0	0	.	0	.	0	.
Eligible for Speech-Language Services	0	0	.	0	.	0	.	1	1	1.00	0	0.00	0	0.00
Multiple Grids and Default	9	7	0.78	2	0.22	0	0.00	8	8	1.00	0	0.00	0	0.00
Section 504	16	7	0.44	8	0.50	1	0.06	99	77	0.78	21	0.21	1	0.01

Part 4: Scoring

4.1 MC items

The answer keys approved by NJ DOE are used to score the multiple-choice items after the responses have been scanned. Each item has a key associated with the item (A, B, C, or D), which has been supplied and verified by the HSPA content specialists. All correct answers are assigned the value of “1” while incorrect answers are assigned the value of “0.” At no time in this process is the original scanned answer overwritten, in case the key is determined to be incorrect during the post-scoring quality assurance check. After scoring is completed, simple item statistics are provided to the appropriate HSPA content specialist to ensure that the correct keys are being applied. If a key changes, then the process is repeated until the scoring file is correct. The key-check data that were provided to the HSPA office for review contains the following information:

- proportion of students getting the question correct (PC);
- correlation of the item to the test as a whole (RPB);
- correlation of each possible response option to the test as a whole (RPBA, RPBB, RPBC, RPBD);
- percentage of students choosing each response option (A, B, C, D or X-omits); and
- flags for items with high difficulty (DifficultyFLAG) or low correlations (PtBis_FLAG).

4.2 OE items

Scorer Selection

Because MI has been conducting the handscoring of writing and open-ended items for many years, a large pool of qualified, experienced readers is available for scoring NJ HSPA. MI routinely maintains supervisors’ evaluations and performance data for each person who works on each scoring project in order to determine employment eligibility for future projects. Many of experienced readers were utilized for scoring NJ HSPA.

New readers are often needed to complement MI’s existing reader pool. MI’s procedures for selecting new readers are very thorough. After advertising in local newspapers, with the job service, and elsewhere, and receiving applications, staff in the human resources department review the applications and schedule interviews for qualified applicants. Qualified applicants are those with a four-year college degree in English, language arts, education, mathematics, science, or a related field. Each qualified applicant must pass an interview by experienced MI staff, write an acceptable essay, and receive good recommendations from references. The information about each applicant is then reviewed before offering employment.

In selecting team leaders, MI’s management staff and scoring directors review the files of all the returning staff. They look for people who are experienced team leaders with a record of good performance on previous projects and also consider readers who have been recommended for promotion to the team leader position.

MI is an equal opportunity employer that actively recruits minority staff. Historically, our temporary staff on major projects averages about 70% female, 30% male, 76% Caucasian and 24% minority. MI strongly opposes illegal discrimination against any employee or applicant for employment with respect to hire, tenure, terms, conditions, or privileges of employment, or any matter directly or indirectly related to employment, because of race, color, religion, sex, age, handicap, national origin, ancestry, veteran status, or sexual orientation.

Field Test Range Finding

Prior to field test scoring, content committees consisting of NJ DOE personnel, NJ teacher representatives, and MI leadership personnel meet in New Jersey to determine “true” scores for 30 selected papers representing each of the score points for each item to be tested. Field test scoring guides and training sets are developed using the papers scored at the range finding.

Operational Range Finding

Range finding meetings are conducted to establish “true” scores for a representative sample of papers. Between 100 and 220 sample papers per task are chosen by MI leadership personnel either from the available field test papers or from the current test administration. For items using specific rubrics, the rubrics are discussed and refined. The sample responses brought to the range finding meetings are selected from a broad range of New Jersey school districts in order to ensure that the sample is representative of overall student performance. The range finding committees consist of NJ DOE content specialists, NJ teacher representatives, MI management personnel, as well as the scoring director responsible for each content area.

Developing Scoring Guides

After the range finding meeting, MI management and the scoring directors develop training materials consisting of an anchor set (examples of responses for each score point) and training/qualifying sets (practice papers) for each task using the responses scored at range finding. Anchor sets usually consist of three, or more, annotated examples of each score point in score point order. Training/qualifying sets consist of clearly anchored papers in random score point order.

Team Leader Training and Qualifying

After the anchor papers, training, and qualifying papers have been identified and finalized, team leader training is conducted by the scoring director for each task, a process which typically takes up to four days depending on the content. Procedures are similar to those for training scorers but are more comprehensive, dealing with resolution of discrepant scores, identification of non-scorable responses, unusual prompt treatment, alert situation responses (e.g., child-in-danger), and other duties performed only by team leaders. Team leaders take careful notes on the training papers in preparation for discussion with the scorers, and the scoring director counsels team leaders on application of the rubric and training techniques. Effective scorer training relies to a

great extent on having knowledgeable, flexible team leaders. Team leaders assist in training scorers in discussions of training sets, and are responsible for distributing, collecting, and accounting for training packets and sample papers during each scoring session. During scoring, team leaders respond to questions, spot-check scorer packets, and counsel scorers having difficulty with the criteria.

Team leaders also administer the quality control validity sets, monitor the scoring patterns of each scorer throughout the project, conduct retraining as necessary, perform some resolution readings, and maintain a professional working environment. Team leaders work 7.75 hours per day.

Scorer Training/Qualifying

All scorers are trained using the rubrics, anchor papers, training papers, and qualifying papers selected during the range finding meetings and approved by the NJ DOE. Scorers are assigned to a scoring group consisting of one team leader and 10-12 scorers. Each scorer is assigned an individual number for easy identification of his or her scoring work throughout the scoring session.

After the contracts and nondisclosure forms are signed, training begins. Scorer training follows the same format as team leader training. The scoring director introduces the set of anchor papers and thoroughly discusses each score point. This presentation is followed by practice scoring on the training sets. Scorers break into teams to discuss the papers in the training sets. This arrangement gives scorers an opportunity to discuss any possible points of confusion or problems in understanding the criteria in a small group setting.

Team leaders collect the monitor sheets after the scoring of each training set, and record results on a customized log which is examined by the scoring director to determine which papers are giving scorers difficulty. The scoring director also “floats” from team to team, listening to the team leaders’ explanations and adding additional information when necessary. If a particular paper or type of paper seems to be causing difficulty across teams, the problem is discussed with the room at large to ensure that everyone hears the same explanation.

Scorers must demonstrate their ability to score accurately by attaining 90% adjacent agreement (within one point) percentage on the qualifying sets before they read packets of actual papers. Any reader unable to meet the standards set by the NJ DOE will be dismissed. All scorers understand this stipulation when they are hired.

Training is carefully orchestrated so that scorers understand how to apply the rubric in scoring the papers, learn how to reference the scoring guide, develop the flexibility needed to deal with a variety of responses, and retain the consistency needed to score all papers accurately. In addition to completing all of the initial training and qualifying, readers are given demonstrations of using the “Virtual Scoring Center” (MI’s image scoring application). This includes instructions on how to send notes with questions about scoring to team leaders and how to identify possible “content concern” alerts as well as instructions about other procedures which are necessary for the conduct of a smooth project. Scorers generally work 7.0 hours per day.

Monitoring Scorer Performance

Each student writing sample will be scored holistically by two independent readers using the Registered Holistic Scoring Method. The two independent scores, if identical or adjacent, will be combined to produce the student's final score on each task. If the two scores differ by more than one score point, the response will be scored by a third reader. The final score is determined by an algorithm supplied by the NJ DOE.

MI project management constantly monitors the quality of each scorer's work throughout every project. The scoring director and team leaders use a variety of methods to insure that the student responses are being scored in accordance with the NJDOE rubrics and directions.

The first line of monitoring is the team leader, who works with each reader on a daily basis. During training readers are reminded that they should refer any questions not addressed in their guides and training sets to their team leader. This allows the team leader to see what papers are causing questions and to make sure that these papers are receiving the correct score. Also, the team leader can use those questions to focus their read-behinds for that reader. Also, if the question is one that the team leader cannot answer, it goes to the scoring director, who can then cover that issue with the entire room.

Team leaders also read behind each reader on a regular basis. Using the VSC system the team leader can set a percentage of each reader's work to be sent automatically to the team leader for review. The percentage can be anything from 0 to 100. The team leader can then identify any response, or type of response, that needs to be discussed with the reader. All such responses are explained to the reader, using guide and training papers as references, and the reader is given the opportunity to correct the score(s). This allows for immediate feedback to readers.

The data application continuously scans the readers' work and generates daily status reports. These reports show the total number of papers read, the number of third readings required, and the percentage agreement of each reader, both perfect and adjacent. The reports also show score point distributions. Scoring directors are experienced in examining the reports and using the information to determine the need for retraining of individual readers or the group as a whole. For items that have a distribution history (equating items), the room distribution can be compared to that of previous years, which allows the scoring director to see if the room as a whole is scoring differently than in previous years. If so, he or she can identify the problem and retrain the entire room on that line. For individual readers, the scoring director and team leaders can see if a particular reader is scoring "high" or "low" according to the room average and, because of the specificity of the reports, they can pinpoint the score point(s) with which the reader is having difficulty. In addition, the VSC system produces a "post-scoring read behind report" which shows exactly how each reader has scored an individual response when compared to the second reader on the same response. This helps focus the team leader's work with that reader. With daily reports, it is possible to see the results of retraining immediately.

In addition to the statistical information from the daily and cumulative reports, there is the validity set report. Each day members of each team take a validity set---a pre-scored and selected set of papers representing the range of scores in each item. The data application records the result and a validity report is printed out daily. Team leaders can check on individual readers

and see on which score points a reader may have problems. Also, the scoring director can see the cumulative responses to each set and see if there is any room-wide issue.

Retraining is an ongoing process once scoring begins. If it becomes apparent that a whole team or a whole group is having difficulty with a particular type of response, large group training sessions are conducted. Standard retraining procedures include room-wide discussions led by the scoring director, team discussions conducted by team leaders, spot-checking of individual scorers by team leaders, and discussions between team leaders and individual scorers.

Scorers are dismissed when, in the opinion of the scoring director and the project director, they have been counseled, retrained, and given every reasonable opportunity to improve and are still performing below the acceptable standard.

4.3 Quality Control Procedures in Data Preparation

Machine Scoring/Linking of Answer Documents

In order to ensure the quality of the testing materials, Measurement Incorporated, the HSPA office, and all sub-contractors work together to rigorously proof all materials prior to printing/production.

As mentioned in Part 2, all items have undergone multiple reviews to ensure that the operational and any field test items being used are valid and fair for all students. All assessment materials are submitted to rigorous editing and proofreading procedures. All copy for materials to be developed are first checked by the editorial staff of Measurement Incorporated prior to being typeset to assure continuity exists across all documents.

Prior to typesetting of any documents, sample layouts based on the approved HSPA Style Guide are provided to HSPA staff for review and approval. Typeset page proofs are then prepared and thoroughly proofread by well-trained staff members who will read the documents in their entirety for typographical errors and for potential problems in context. Copies of the page proofs are provided to the HSPA coordinator at NJ DOE for review and approval. Upon approval of the page proofs, printer's proofs are produced. Printer's proofs of all documents are proofread by two staff members and two independent editors and then provided to the HSPA coordinator for final approval prior to printing. All forms necessary for test administration are approved by NJ DOE prior to final production.

Additionally, all accommodated materials are also reviewed for accuracy and quality at multiple stages. The first stage of review involves content specialist ensuring that the items being used on the tests are still valid in the accommodated format. Once this is completed, the large print test follows the quality control procedures discussed previously, while the other formats undergo additional quality control procedures.

Part 5: Item-Level Statistics

5.1 Classical Item Statistics

Tables 5.1.1 and 5.1.2 present the item analysis summary for Language Arts Literacy and Mathematics, respectively. The item analysis provides statistical information about the items that helps identify items that may be mis-keyed. The following information is presented in each item analysis:

Classical Item Difficulties: The mean item score is used to identify items that are potentially too difficult or too easy. In the case of multiple-choice items, the mean item score is the same as the proportion of students who answer each item correctly (p -value). Generally, multiple-choice items are selected so that the p -values are greater than 0.25 and less than 0.95. Any mis-keyed items are corrected and the test is rescored.

Item Discriminations: The point biserial correlation and item-total correlation are presented as a measure of item discrimination. The item discrimination measures the relationship between the item score and the total score. The higher the correlation the better the item discriminates. The point biserial for the keyed response is presented for multiple-choice items. The item-total correlation is presented for the open-ended items.

Table 5.1.1: Item Difficulty and Discrimination Summary Statistics for Multiple-Choice and Open-Ended items by Content Area and Cluster (Standard)

Test Section/Cluster (Standard)	Multiple-Choice				Open-Ended			
	Item Difficulty			Item Discrimination	Item Difficulty			Item Discrimination
	Number of Items	Mean	S.D.	Mean	Number of Items	Mean	S.D.	Mean
LAL	20	0.36	0.07	0.25	6	1.53	1.43	0.63
Writing								
Expository					1	2.02	.	0.65
Persuasive					1	4.25	.	0.78
Reading	20	0.36	0.07	0.25				
Interpreting Text	10	0.34	0.08	0.23				
Analyzing Text	10	0.39	0.06	0.28	4	0.73	0.07	0.58
Math								
Math	30	0.39	0.19	0.16	6	0.47	0.16	0.51
Number & Numerical Operations	4	0.42	0.32	0.18	1	0.43	.	0.44
Geometry and Measurement	6	0.36	0.13	0.14	2	0.44	0.13	0.50
Patterns and Algebra	9	0.43	0.20	0.14	2	0.64	0.10	0.57
Data Analysis, Probability, and Discrete Mathematics	11	0.37	0.18	0.17	1	0.25	.	0.46
Problem Solving	24	0.33	0.15	0.15	6	0.47	0.16	0.51

Table 5.1.2: Frequency Distribution for Multiple-Choice P-values and Discrimination Indices by Content Area and Cluster (Standard)

	Number of Items	P-value							Discrimination				
		Median	$P < 0.25$	$0.25 \leq p < 0.40$	$0.40 \leq p < 0.55$	$0.55 \leq p < 0.70$	$0.70 \leq p < 0.85$	$P \geq 0.85$	Median	$pb < 0.25$	$0.25 \leq pb < 0.35$	$0.35 \leq pb < 0.45$	$0.45 \leq pb < 0.55$
LAL (Reading)	20	0.34	.	14	6	.	.	.	0.25	10	7	2	1
Interpreting Text	10	0.32	.	8	2	.	.	.	0.23	7	2	.	1
Analyzing Text	10	0.38	.	6	4	.	.	.	0.26	3	5	2	.
Math	30	0.35	8	13	1	6	2	.	0.15	27	3	.	.
Number & Numerical Operations	4	0.38	2	.	.	1	1	.	0.19	4	.	.	.
Geometry and Measurement	6	0.34	2	2	1	1	.	.	0.12	5	1	.	.
Patterns and Algebra	9	0.36	2	4	.	2	1	.	0.14	8	1	.	.
Data Analysis, Probability, and Discrete Mathematics	11	0.36	2	7	.	2	.	.	0.16	10	1	.	.
Problem Solving	24	0.33	8	12	1	3	.	.	0.15	21	3	.	.

5.2 Speededness

The NJ HSPA is intended to provide sufficient time for almost all students to respond to all of the items. The percentage of students omitting an item provides information about speededness. If the percentage is low, it implies that speededness is probably not an issue. If the percentage of omits is high, speededness may be a problem, although other factors may be contributing to students omitting items besides not having enough time.

Table 5.2.1 presents data regarding omit rates.

Table 5.2.1: NJ HSPA Percentage of Students Omitting the Last Item of Each Part

Subject	Multiple-Choice		Open-Ended	
	Order	Omitting %	Order	Omitting %
LAL	9	3.11	11	19.8
	10	3.20	12	26.2
	21	6.89	23	18.9
	22	7.31	24	21.2
Math	09	1.06	11	8.60
	10	0.69	12	16.7
	21	0.37	23	11.5
	22	0.55	24	9.69
	33	0.93	35	14.2
	34	2.33	36	22.4

5.3 Intercorrelations

The Pearson product-moment correlations between the content areas and test sections/clusters are presented in Table 5.3.1. Generally, if a cluster (standard) has more items then its correlation with the total score will be higher. After all, the cluster (standard) makes up more of the points of the total score. For example, the Reading total score is highly correlated with the Language Arts Literacy (LAL) score (.95) because the Reading score makes up 36 out of the 54 possible points for LAL. For Mathematics, the correlation between the Mathematics Multiple-Choice (M MC) total and the total Mathematics score (MAT) is 0.95. This is partly due to the fact that M MC is 30 of the 48 total points for Mathematics.

Table 5.3.2 shows the correlations between students' scores and content clusters (standards). As was true with table 5.3.1, the correlations between Mathematics standard scores and Language Arts Literacy cluster scores tended to be low.

Table 5.3.1: NJ HSPA Intercorrelations Among Major Content Clusters (Standards) and Item Types

Major Content Clusters and Item Types	Major Content Clusters (Standard) and Item Types								
	Language Arts Literacy (LAL)						Mathematics (MAT)		
	LAL	R	R MC	R OE	R P1	R P2	MAT	M MC	M OE
LAL Language Arts Literacy (54)	1.00	0.95	0.87	0.78	0.79	0.86	0.57	0.48	0.59
R Reading (36)	0.95	1.00	0.93	0.78	0.86	0.88	0.55	0.47	0.57
R MC Reading Multiple-Choice (20)	0.87	0.93	1.00	0.51	0.80	0.82	0.51	0.44	0.51
R OE Reading Open-Ended (16)	0.78	0.78	0.51	1.00	0.67	0.69	0.45	0.37	0.49
R P1 Reading Passage 1 (18)	0.79	0.86	0.80	0.67	1.00	0.51	0.49	0.41	0.51
R P2 Reading Passage 2 (18)	0.86	0.88	0.82	0.69	0.51	1.00	0.47	0.40	0.48
MAT Mathematics (48)	0.57	0.55	0.51	0.45	0.49	0.47	1.00	0.95	0.88
M MC Mathematics Multiple-Choice (30)	0.48	0.47	0.44	0.37	0.41	0.40	0.95	1.00	0.68
M OE Mathematics Open-Ended (18)	0.59	0.57	0.51	0.49	0.51	0.48	0.88	0.68	1.00

Table 5.3.2: NJ HSPA Intercorrelations Among Content Areas and Clusters (Standards)

Test Section/ Cluster (Standard)	Test Section/ Cluster (Standard)											
	LAL Language Arts Literacy							MAT Mathematics				
	LAL	L1	L2	L3	L4	L5	L6	MAT	M1	M2	M3	M4
LAL Language Arts Literacy (54)	1.00	0.95	0.73	0.91	0.83	0.66	0.78	0.57	0.40	0.46	0.52	0.44
L1 Reading (36)	0.95	1.00	0.79	0.94	0.62	0.51	0.57	0.55	0.39	0.46	0.49	0.43
L2 Interpreting Text (10)	0.73	0.79	1.00	0.54	0.42	0.35	0.38	0.47	0.32	0.38	0.43	0.36
L3 Analyzing Text (26)	0.91	0.94	0.54	1.00	0.62	0.51	0.58	0.51	0.36	0.42	0.45	0.39
L4 Writing (18)	0.83	0.62	0.42	0.62	1.00	0.77	0.95	0.45	0.33	0.35	0.42	0.34
L5 Writing Expository (6)	0.66	0.51	0.35	0.51	0.77	1.00	0.54	0.42	0.29	0.32	0.40	0.33
L6 Writing Persuasive (12)	0.78	0.57	0.38	0.58	0.95	0.54	1.00	0.40	0.29	0.31	0.36	0.30
MAT Mathematics (48)	0.57	0.55	0.47	0.51	0.45	0.42	0.40	1.00	0.71	0.81	0.86	0.81
M1 Number and Numerical Operations (7)	0.40	0.39	0.32	0.36	0.33	0.29	0.29	0.71	1.00	0.50	0.51	0.47
M2 Geometry and Measurement (12)	0.46	0.46	0.38	0.42	0.35	0.32	0.31	0.81	0.50	1.00	0.58	0.52
M3 Patterns and Algebra (15)	0.52	0.49	0.43	0.45	0.42	0.40	0.36	0.86	0.51	0.58	1.00	0.56
M4 Data Analysis, Probability, and Discrete Math (14)	0.44	0.43	0.36	0.39	0.34	0.33	0.30	0.81	0.47	0.52	0.56	1.00

Part 6: Standard Setting

In Spring of 2002, standard setting for the Mathematics and Language Arts Literacy tests took place. As for other NJ assessments, the results of standard setting (raw cut scores at Proficient and Advanced Proficient) were used to generate a linear raw score to scaled score line with Proficient set at 200 and Advanced Proficient set at 250. For these tests, the following cut scores were set by NJ educators and approved by the state board:

Table 6.1.1: Target HSPA Cut Scores and Rasch thetas from Spring 2002 Standard Setting

Levels	Math		LAL	
	Raw	Theta	Raw	Theta
Proficient	22.5	0.055	29.5	0.2815
Advanced	39.0	1.706	42.0	2.6980

Using the Rasch model, thetas representing student ability were identified at the following points. Additional explanation of the Rasch model, theta and equating can be found in the next section. The following raw score cuts were obtained as the result of equating procedures described in Part 7 of this report.

Table 6.1.2: Obtained HSPA Math Spring Cut Scores

Levels	2002	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Proficient	22.5	20.0	21.5	20.5	24.0	19.5	18.5	17.5	17.5	17.0	16.5	18.0	19.5
Advanced	39.0	35.0	38.0	37.5	40.5	35.5	35.5	37.0	36.5	35.5	34.5	35.5	36.5
Total Points	47*	47*	48	48	48	48	48	48	48	48	48	48	48

*One item dropped

Table 6.1.3: Obtained HSPA LAL Spring Cut Scores

Levels	2002	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Proficient	29.5	27.5	24.0	24.0	25.5	29.0	25.5	22.5	20.0	20.5	19.5	19.0	19.5
Advanced	42.0	40.5	38.5	38.0	38.5	41.5	39.5	37.5	35.5	35.5	35.5	35.5	36.0
Total Points	54	54	54	54	54	54	54	54	54	54	54	54	54

The Spring 2015 raw cut for LAL was 0.5 points higher for both Proficient and Advanced when compared to the Spring 2014 test. For Mathematics, the raw cut was 1.5 points higher for Proficient and 1.0 point higher for Advanced when compared to the Spring 2014 test. In identifying the appropriate cut, the exact theta from standard setting is seldom obtained. Thus, the following shows the obtained theta at cut. The executive summary of the Spring 2002 Math and LAL Standard Setting can be found in Appendix E.

Table 6.1.4: Obtained Math Rasch thetas for the raw Cut Scores in Table 6.1.2

Levels	2002	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Proficient	0.055	0.030	0.059	0.052	0.042	0.032	0.063	0.053	0.048	0.043	0.055	0.024	0.033
Advanced	1.706	1.701	1.678	1.692	1.708	1.675	1.693	1.726	1.714	1.701	1.710	1.698	1.717

Table 6.1.5: Obtained LAL Rasch thetas for the raw Cut Scores in Table 6.1.3

Levels	2002	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Proficient	0.282	0.234	0.266	0.287	0.289	0.279	0.293	0.276	0.268	0.296	0.283	0.242	0.256
Advanced	2.698	2.657	2.676	2.635	2.644	2.667	2.692	2.697	2.663	2.656	2.718	2.638	2.695

Part 7: Scaling and Equating

7.1 Scaling

The total scores in the NJ HSPA Language Arts Literacy and Mathematics test are reported as scale scores with a theoretical range of 100 to 300. The scale score of 200 is the cut point between Partially Proficient and Proficient students. The scale score of 250 is the cut point between Proficient and Advanced Proficient students. The score range follows:

Partially Proficient	100 – 199
Proficient	200 – 249
Advanced Proficient	250 – 300

The students who score above 200, that is, Proficient or Advanced Proficient, are above the state minimum level of proficiency. Any student who scores below 200, Partially Proficient, has at least 2 additional opportunities to pass the NJ HSPA.

Manual Adjustments to Scaled Score Lookup Tables

The scaled score lookup tables for Math and LAL are based on an interpolation back to the Spring 2002 scaled score table using the estimated raw score (ERS) and the slope and intercept parameter established in the first administration of the HSPA exams in Spring 2002. The slope and intercept parameters for the Language Arts Literacy exam are 4 and 82, respectively. The slope and intercept parameters for the Mathematics exam are 3.03 and 131.83, respectively.

NJ DOE policy requires that scaled scores below 100 be rounded up to 100 and scaled scores above 300 are rounded down to 300. For all regular tests, NJ also requires perfect scores set at 300. All tests required that 300 be set manually for perfect scores. Per NJ DOE policy, where there is a score of below and above 200, but not exactly 200, the scaled score under 200 should be adjusted upward to 200. Additional details on the Spring 2015 Math and LAL Equating may be found in the “Final Spring 2015 Equating Report” generated by Measurement Incorporated for NJ DOE.

7.2 Equating

Description of Equating Procedure

These tests are equated using a Rasch model to estimate the latent trait difficulty of each item on the test. The software used to do this is called WINSTEPS. WINSTEPS is a program written by Linacre (Linacre, 2004) to calculate Rasch analyses. The program is a WINDOWS-based program. It is widely used for similar high stakes tests. WINSTEPS (the Rasch model) allows for the estimation of item difficulty for multiple-choice and open-ended items on a single scale. Using these item difficulties, the model is able to estimate the necessary ability (theta) of a student to earn each raw score on the test.

To equate this year’s tests, a raw to Rasch table was generated for each assessment as has been done with other assessments. Note that, consistent with other New Jersey assessments, data

entered into WINSTEPS had to be adjusted to accommodate the use of half-points and one Writing prompt that is the sum of two readers' scores. Use of these half-point increments is problematic for the WINSTEPS software used. The second part of this approach requires double weighting each item using the IWEIGHT option in WINSTEPS. To accommodate the half-point scoring for open-ended items in Mathematics and Language Arts Literacy, the rater one score is used instead of the final (average of reader one and reader two) score. After a raw score to Rasch table is generated (Table 20.1 in WINSTEPS), each raw score is divided by 2 to re-create the half-point steps obtainable on the test. These steps are done for equating purposes only.

Language Arts Literacy Equating

In Spring of 2002, data from about 75,000 students – what is called the general population – was used for impact data for standard setting. Due to NJ HSPA 2015 assessing only students who were retesting, the selection of students selected to represent the different DFGs in New Jersey was unable to represent the general population. For Language Arts Literacy, only 479 students were selected as having valid scores for equating. Below are the percent of each gender and ethnicity from the Spring 2002 population data, and the equating sample from the last five years. As can be seen in Table 7.2.1, the sample chosen for the 2015 equating is not representative of the population used in standard setting.

Table 7.2.1: Sample Demographics for LAL 2002, 2011-2015 General Population

GENDER	Sp 02 Percent	Sp 11 Percent	Sp 12 Percent	Sp 13 Percent	Sp 14 Percent	Sp 15 Percent
F	50.4	52.8	52.6	53.0	52.1	44.7
M	49.6	47.2	47.4	47.0	47.9	55.3
Ethnic	Sp 02 Percent	Sp 11 Percent	Sp 12 Percent	Sp 13 Percent	Sp 14 Percent	Sp 15 Percent
Asian	6.9	11.2	10.3	14.0	10.6	4.4
Black	13.5	15.7	11.1	16.3	14.9	38.4
Hispanic	11.7	14.7	17.8	18.8	18.2	37.8
Indian	0.2	0.2	0.3	0.2	0.3	0.2
Pacific Islander	0.4	0.2	0.4	0.3	0.2	0.2
White	65.9	58.3	59.0	51.4	55.8	14.4

Historically, for the Language Arts Literacy test equating, a process called forward equating was used. Given the small and unrepresentative sample size the items were equated in a different manner. Items difficulty parameters were linked back to previous test administrations. For Language Arts Literacy, all of the items associated with the two base test passages had been in embedded field test positions on the Spring 2013 tests. Those items were anchored to the Spring 2013 base test to be on the same scale. The set consisted of 20 multiple-choice and four 4-pt. constructed-response questions – all of the items on the test except for the expository and persuasive writing prompt. The item-difficulty and step parameters of the writing tasks were linked back to the years they were field tested.

LAL Item Evaluation Process

It should be noted that other analyses are done to guarantee the appropriateness of items for use in generating student test scores. An item analysis was conducted on multiple-choice items to validate the keys. Part 5.1 includes the item analyses for the multiple-choice questions. All multiple-choice items were determined to have been scored correctly.

In previous years, the LAL anchor item values obtained from previous test administrations were compared with their values to this year's test. Items common to the two forms served as anchors. This evaluation determines which items are stable enough to be considered as anchors in later analyses. Given the small and unrepresentative samples the items were evaluated this year using a different procedure. All items were anchored to previous test administrations. The anchors runs were conducted and the infit and outfit measures were evaluated for each of the items. Fit indices are a Rasch-based measure of the degree to which the model can summarize the items and the students using the data provided. A large number of items with infit or outfit outside of the bounds of .7 to 1.3 can indicate problems in the item set (outfit) or unusual student performance (infit). Items that were considered misfitting were unanchored until model fit was within reasonable limits. For LAL, given the extremely small sample size, the limits of infit and outfit were increased to .6 and 1.4. Refer to the 2015 NJ HSPA Equating Report for a more detailed explanation of the equating process.

After the anchor evaluation process, a second examination of item fit statistics from the regular administrations was conducted. The data below summarize the infit and outfit data from the regular administrations after the anchor evaluation process was conducted. It should be noted that infit and outfit statistics have a center at 1.0.

Table 7.2.3: Language Arts Literacy Infit/Outfit Summary Statistics

Infit Summary

N	Mean	Std Dev	Minimum	Maximum
26	1.05	0.19	0.56	1.39

Outfit Summary

N	Mean	Std Dev	Minimum	Maximum
26	1.10	0.20	0.83	1.45

There were five LAL item with infit statistics outside of the range of 0.7 to 1.3; they had an infit values of 0.56, 1.31, 1.37, 1.38, and 1.39. There were seven items with outfit statistics outside of the range; their outfit values were 1.31, 1.31, 1.33, 1.37, 1.38, 1.42 and 1.45. This suggests that the model had difficulty with some of the items; that is likely due to the relatively small and skewed sample size.

Mathematics Equating

In Spring of 2002, data from about 75,000 students – what is called the general population – was used for impact data for standard setting. Due to NJ HSPA 2015 assessing only students who were retesting, the selection of students selected to represent the different DFGs in New Jersey was unable to represent the general population. For mathematics, only 3791 students were selected as having valid scores for equating. Below are the percent of each gender and ethnicity from the Spring 2002 population data, and the equating sample from the last five years. As can be seen in Table 7.2.4, the sample chosen for this equating is not representative of the population used in standard setting.

Table 7.2.4: Sample Demographics for MATH 2002, 2011-2015 General Population

GENDER	Sp 02 Percent	Sp 11 Percent	Sp 12 Percent	Sp 13 Percent	Sp 14 Percent	Sp 15 Percent
F	50.4	52.7	52.4	53.1	52.1	59.1
M	49.6	47.3	47.6	46.9	47.9	40.9
Ethnic	Sp 02 Percent	Sp 11 Percent	Sp 12 Percent	Sp 13 Percent	Sp 14 Percent	Sp 15 Percent
Asian	6.9	10.7	10.2	13.7	10.6	2.8
Black	13.5	15.4	11.1	16.2	15.0	41.4
Hispanic	11.7	15.2	17.8	18.8	18.3	36.9
Indian	0.2	0.2	0.3	0.2	0.2	0.1
Pacific Islander	0.4	0.2	0.4	0.3	0.2	0.3
White	65.8	58.5	58.9	51.8	55.8	16.5

For the Mathematics test, 17 multiple-choice and three 3-pt. constructed-response questions were initially identified as potential anchors. Those items were anchored to the Spring 2014 base test to be on the same scale. The other 13 multiple-choice and three constructed-response items were linked back to the years they were field tested.

Math Item Evaluation Process

It should be noted that other analyses are done to guarantee the appropriateness of items for use in generating student test scores. An item analysis was conducted on multiple-choice items to validate the keys. Part 5.1 includes the item analyses for the multiple-choice questions. All multiple-choice items were determined to have been scored correctly.

In previous years, the math anchor item values obtained from previous test administrations were compared with their values to this year's test. Items common to the two forms served as anchors. This evaluation determines which items are stable enough to be considered as anchors in later analyses. Given the small and unrepresentative samples the items were evaluated this year using a different procedure. All items were anchored to previous test administrations. The anchors runs were conducted and the infit and outfit measures were evaluated for each of the items. Fit indices are a Rasch-based measure of the degree to which the model can summarize the items and the students using the data provided. A large number of items with infit or outfit outside of the bounds of .7 to 1.3 can indicate problems in the item set (outfit) or unusual student performance (infit). Items that were considered misfitting were unanchored until model fit was within

reasonable limits. Refer to the 2015 NJ HSPA Equating Report for a more detailed explanation of the equating process.

After the anchor evaluation process, a second examination of item fit statistics from the regular administrations was conducted. The data below summarize the infit and outfit data from the regular administrations after the anchor evaluation process was conducted. It should be noted that infit and outfit statistics have a center at 1.0; the mathematics infit and outfit statistics are centered almost perfectly.

Table 7.2.6: Mathematics Infit/Outfit Summary Statistics

Infit Summary

N	Mean	Std Dev	Minimum	Maximum
36	1.00	0.10	0.76	1.22

Outfit Summary

N	Mean	Std Dev	Minimum	Maximum
36	1.01	0.12	0.75	1.27

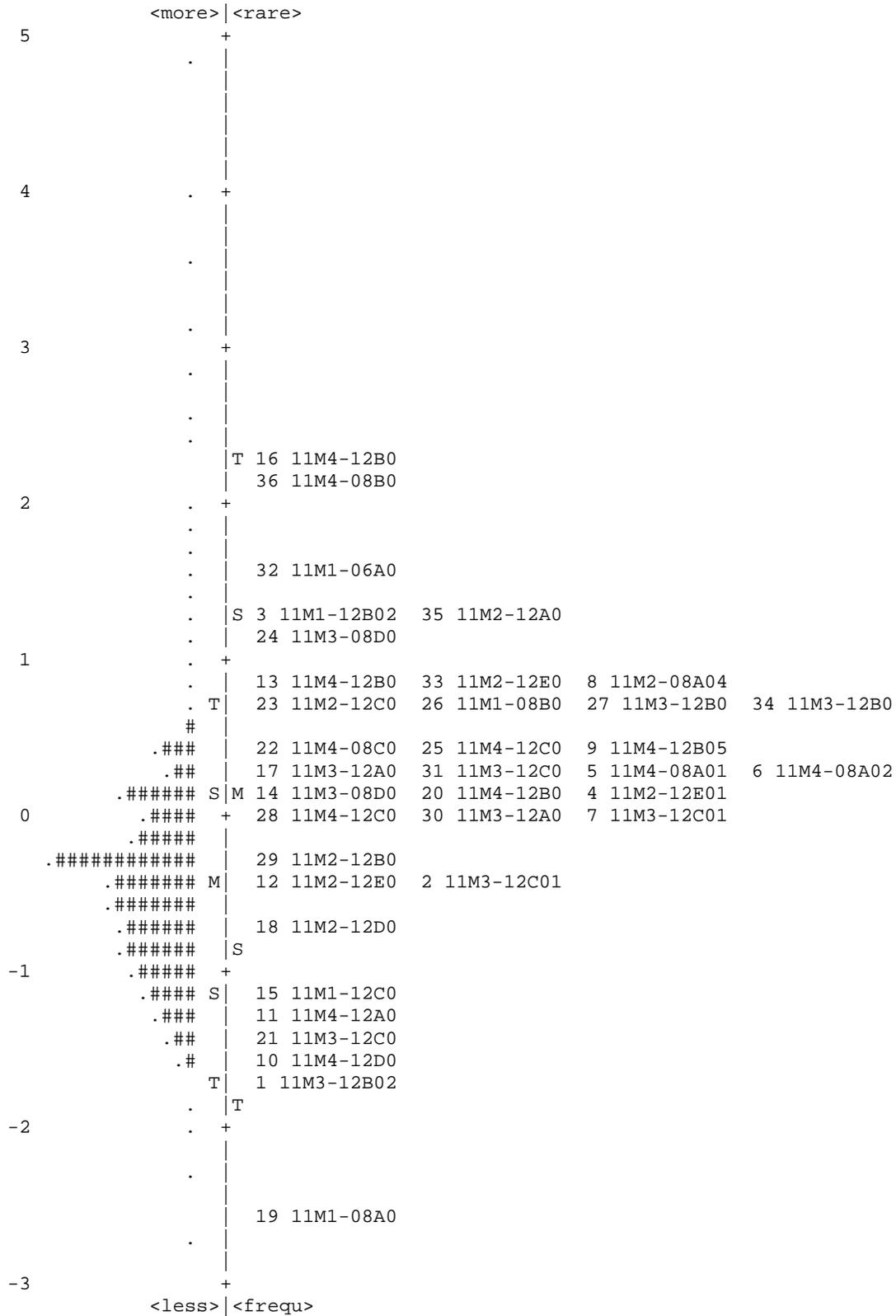
There were no Math items with infit or outfit values outside of the range of .7 to 1.3. Furthermore, the mean infit and outfit values were extremely close to 1.00. There is no indication of a problem with the data fitting the Rasch model.

Item Maps and Test Information Functions

Item maps for LAL and Math are presented in figures 7.2.1 – 7.2.2. These Figures indicate how well the item difficulties and person ability levels match.

Figure 7.2.2 Math Item Map

PERSONS MAP OF ITEMS



EACH '#' IS 44.

Figure 7.2.3 LAL Test Information Function Figure

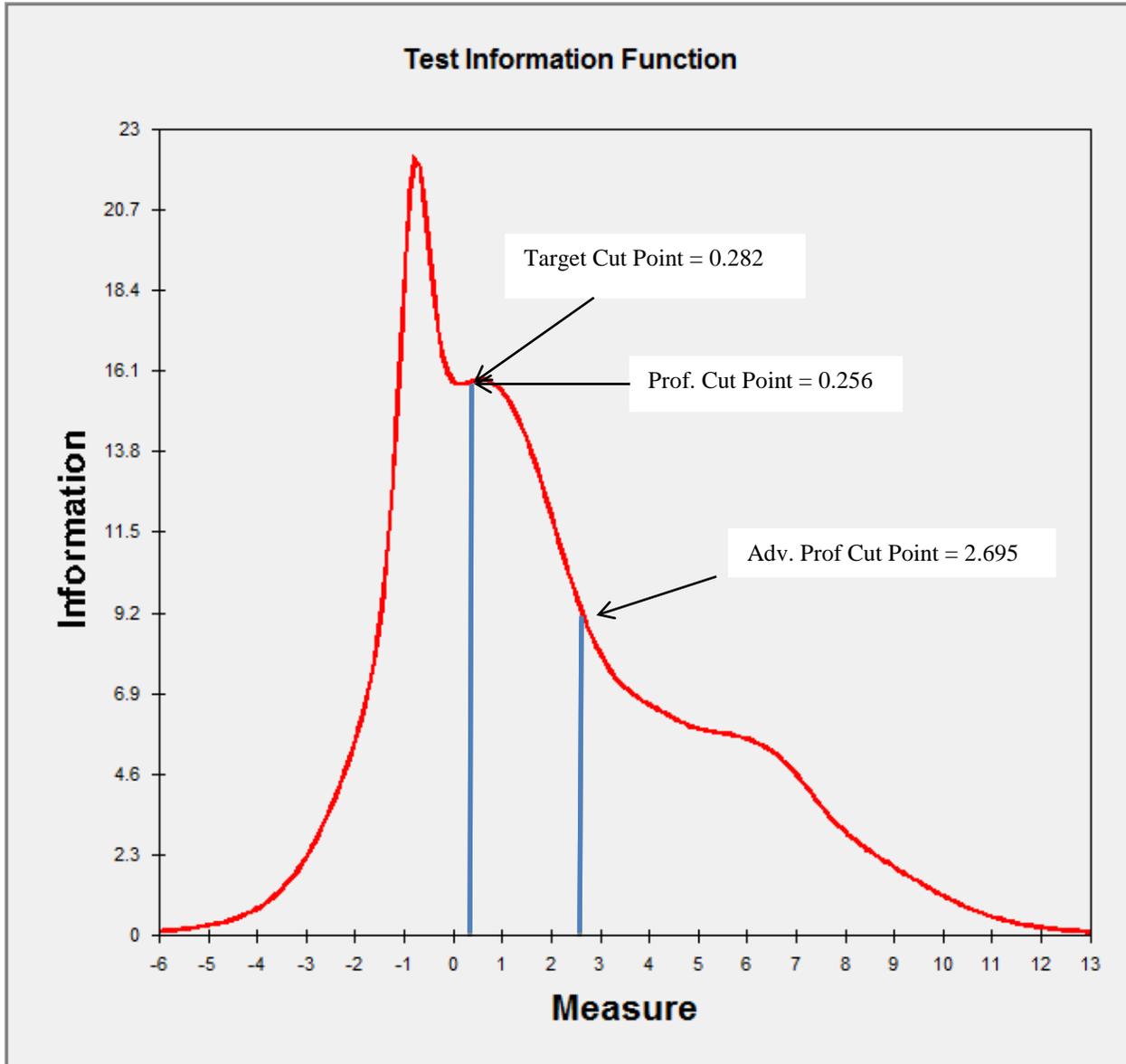
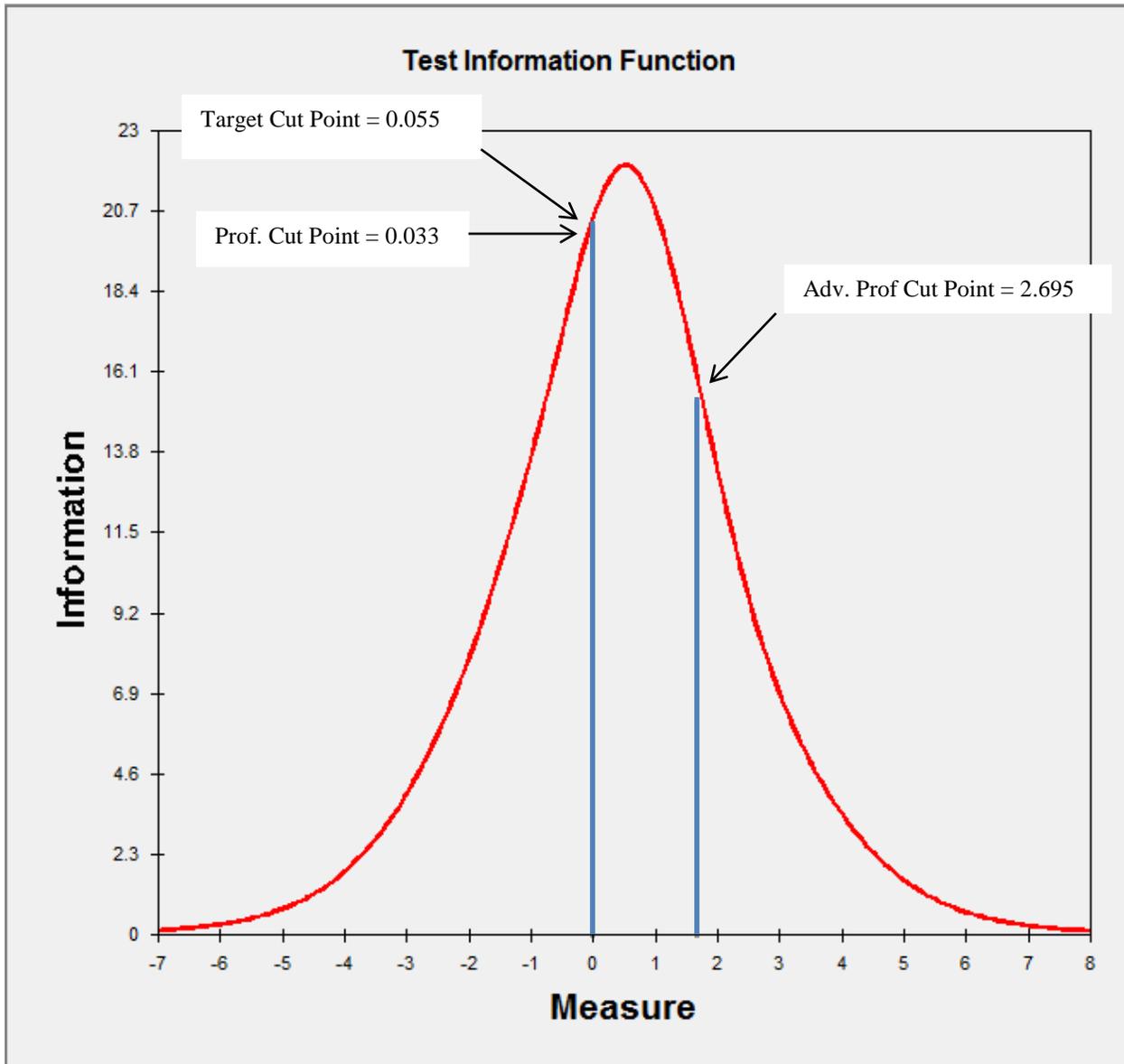


Figure 7.2.4 Math Test Information Function Figure



Braille, Alternate and Large Print Equating

During previous administrations, Braille and Large Print versions of the assessment might have required a “special equating.” In 2015 the Braille and Large Print assessments consisted of all the same items as the regular test administration. Thus, no “special equating” was required. Furthermore, there was no need to administer the alternate LAL or math forms because there were no test breaches. Thus, no alternate equating was required. Additional details on the Spring 2015 Equating may be found in the “Final Spring 2015 Equating Report” generated by Measurement Incorporated for NJ DOE.

Open-Ended Scoring Summary

Table 7.2.3 and Table 7.2.4 show the distribution of open-ended points for LAL and Math, respectively. The percentage of students at each score point is listed below each possible score.

Table 7.2.7: Language Arts Literacy Open-Ended Score Distribution

Code	Points	Rasch	Mean	SD	Corr	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6
OE1	4	3.3644	0.81	0.71	0.60	1200	334	1157	502	352	34	18	1	1	0	0	0	0
OE2	4	3.4863	0.66	0.66	0.53	1455	460	1061	379	200	30	12	0	2	0	0	0	0
OE3	4	3.4034	0.69	0.65	0.58	1303	498	1198	355	200	30	14	1	0	0	0	0	0
OE4	4	3.1916	0.77	0.71	0.63	1250	504	991	464	320	50	19	0	1	0	0	0	0
WR1	6	2.4294	2.02	0.94	0.65	321	0	388	303	1241	544	582	130	74	11	3	1	1
Code	Points	Rasch	Mean	SD	Corr	0	1	2	3	4	5	6	7	8	9	10	11	12
WR2	12	1.5873	4.25	2.04	0.78	386	0	297	234	978	595	795	196	101	13	2	2	0

Table 7.2.8: Mathematics Open-Ended Score Distribution

Code	Points	Rasch	Mean	SD	Corr	0.0	0.5	1.0	1.5	2.0	2.5	3.0
OE1	3	0.2840	0.71	0.94	0.57	6024	572	2003	191	1431	187	640
OE2	3	1.5707	0.43	0.63	0.44	6905	403	2877	116	690	38	19
OE3	3	0.9186	0.35	0.60	0.50	7649	275	2681	90	130	38	185
OE4	3	0.7573	0.57	0.78	0.57	6359	567	2236	470	1084	122	210
OE5	3	1.2942	0.53	0.61	0.50	5719	303	4321	79	592	6	28
OE6	3	2.0803	0.25	0.49	0.46	8496	312	1815	164	241	5	15

Part 8: Test Statistics

8.1 Summary Statistics

Mean and standard deviation of students' raw scores on each content area are given in Table 8.1.1 for the March 2015 test. Raw score to scale score conversion tables and frequency distributions are included in Appendix B.

Table 8.1.1: NJ HSPA Raw Score Means and Standard Deviations

Test	Number of Points	Raw Score Mean	Standard Deviation	Number Tested
Language Arts Literacy	54	16.48	6.84	3599
Mathematics	48	14.56	5.31	11048

Table 8.1.2 reports the mean and standard deviation for students' raw scores by cluster (standard) for the Spring 2015 tests. The number of raw score points for both multiple-choice and open-ended are presented. Additionally, the mean percent correct score for each cluster (standard) is also presented.

Table 8.1.2: 2015 NJ HSPA Means and Standard Deviations of Students' Raw Score and Percent Correct by Cluster (Standard)

NJ HSPA Content Area	Number of Items		Number of Possible Points	Raw Score		Percent Correct	
	Multiple-Choice	Open-Ended		Mean	STD	Mean	STD
Language Arts Literacy	20	6	54	16.48	6.84	30.52	12.66
Writing	0	2	18	6.27	2.66	34.81	14.75
Writing/Expository	0	1	6	2.02	0.94	33.61	15.66
Writing/Persuasive	0	1	12	4.25	2.04	35.41	17.03
Reading	20	4	36	10.22	4.87	28.38	13.54
Interpreting Text	10	0	10	3.35	1.94	33.52	19.36
Analyzing Text	10	4	26	6.86	3.56	26.40	13.68
Mathematics	30	6	48	14.56	5.31	30.34	11.07
Number and Numerical Operations	4	1	7	2.13	1.14	30.43	16.27
Geometry and Measurement	6	2	12	3.02	1.73	25.15	14.39
Patterns and Algebra	9	2	15	5.11	2.36	34.05	15.76
Data Analysis, Probability, and Discrete Math	11	1	14	4.31	1.97	30.77	14.07

8.2 Classical Reliability Estimates of the Test Scores

Table 8.2.1 summarizes reliability estimates for HSPA content areas and clusters (standards). The reliability coefficients are based on Cronbach's alpha measure of internal consistency. The standard errors of measurement (SEMs) for the major content areas are expressed in terms of the raw score metric and the scale score metric. When evaluating these results it is important to

remember that reliability is partially a function of test length and thus reliability is likely to be greater for clusters that have more items. Furthermore, it is also a function of the homogeneity of the ability levels of the examinees. The more heterogeneous the ability level of the examinees the more likely the reliability level of the test will increase. The low reliability estimates for some of the clusters are likely due to having fewer items and a relatively homogeneous population of examinees.

Table 8.2.1: 2012 NJ HSPA Reliability Estimates and Standard Errors of Measurement for Content Areas and Clusters (Standards)

NJ HSPA Test Section	Number of Points	Reliability	Raw Score SEM	Scale Score SEM
Language Arts Literacy	54	0.79	3.17	12.36
Writing	18	0.57	.	.
Reading	36	0.76	.	.
Interpreting Text	10	0.47	.	.
Analyzing Text	26	0.72	.	.
Mathematics	48	0.70	2.90	10.20
Number and Numerical Operations	7	0.23	.	.
Geometry and Measurement	12	0.36	.	.
Patterns and Algebra	15	0.43	.	.
Data Analysis, Probability, and Discrete Math	14	0.39	.	.

Subgroups of interest, as requested by NJ DOE were analyzed for their reliability as well. Table 8.2.2 contains the reliability estimates and standard errors of measurement in LAL for subgroups requested at the whole test level as well as at the strand level. Table 8.2.3 contains the reliability estimates and standard errors of measurement in Mathematics for subgroups requested at the whole test level as well as at the strand level.

Table 8.2.2: NJ HSPA Reliability Estimates and Standard Errors of Measurement for Content Areas and Clusters (Standards) for Subgroups in Language Arts

Subgroup*	N	Mean	Test Level			Reading			Writing			Interpreting Text			Analyzing Text		
			Rel.	SD	SEM	Rel.	SD	RS SEM	Rel.	SD	RS SEM	Rel.	SD	RS SEM	Rel.	SD	RS SEM
			General Education	3113	16.66	0.79	6.90	3.17	0.76	4.91	2.38	0.58	2.67	1.74	0.47	1.94	1.42
Special Education	486	15.37	0.76	6.34	3.11	0.74	4.58	2.32	0.50	2.53	1.79	0.44	1.88	1.40	0.71	3.39	1.82
Limited English Proficient	1305	14.39	0.66	5.12	3.00	0.56	3.56	2.35	0.65	2.39	1.41	0.14	1.54	1.42	0.58	2.85	1.86
Male	2056	16.09	0.77	6.54	3.15	0.74	4.66	2.36	0.55	2.61	1.74	0.44	1.90	1.42	0.69	3.38	1.87
Female	1513	17.00	0.80	7.19	3.19	0.78	5.13	2.38	0.58	2.72	1.76	0.50	1.98	1.41	0.74	3.75	1.90
White	392	20.31	0.85	8.24	3.21	0.85	6.07	2.39	0.62	2.82	1.74	0.58	2.23	1.44	0.81	4.35	1.88
Black or African American	1142	16.35	0.76	6.48	3.14	0.75	4.68	2.34	0.41	2.50	1.92	0.47	1.92	1.39	0.70	3.41	1.86
Asian	188	18.12	0.74	5.93	3.00	0.70	4.36	2.40	0.59	2.29	1.46	0.41	1.86	1.43	0.67	3.32	1.90
Hispanic or Latino	1707	15.25	0.74	6.11	3.12	0.69	4.23	2.37	0.60	2.59	1.64	0.34	1.75	1.42	0.65	3.17	1.88
Economically Disadvantaged	2178	15.51	0.72	5.94	3.12	0.68	4.15	2.36	0.58	2.56	1.65	0.34	1.74	1.41	0.64	3.13	1.87
Non-Economically Disadvantaged	1421	17.97	0.83	7.79	3.23	0.82	5.65	2.39	0.53	2.74	1.88	0.57	2.15	1.41	0.77	4.03	1.91

* Due to small sample sizes the Native Hawaiian or Other Pacific Islanders and American Indian or Alaska Native categories have been suppressed.

Table 8.2.3: NJ HSPA Reliability Estimates and Standard Errors of Measurement for Content Areas and Clusters (Standards) for Subgroups in Mathematics

Subgroup	N	Mean	Test Level			Number and Numerical Operations			Geometry and Measurement			Patterns and Algebra			Data Analysis, Probability, and Discrete Math		
			Rel.	SD	SEM	Rel.	SD	RS SEM	Rel.	SD	RS SEM	Rel.	SD	RS SEM	Rel.	SD	RS SEM
General Education	10010	14.69	0.70	5.31	2.90	0.23	1.13	0.99	0.37	1.74	1.38	0.43	2.36	1.78	0.39	1.97	1.54
Special Education	1038	13.37	0.70	5.21	2.86	0.19	1.18	1.06	0.27	1.55	1.33	0.44	2.32	1.73	0.40	1.97	1.52
Limited English Proficient	1699	12.32	0.68	4.81	2.73	0.19	1.06	0.95	0.38	1.68	1.32	0.44	2.18	1.62	0.28	1.73	1.47
Male	4971	14.74	0.73	5.55	2.91	0.26	1.18	1.02	0.39	1.74	1.36	0.45	2.42	1.79	0.42	2.03	1.55
Female	6029	14.43	0.68	5.10	2.89	0.20	1.10	0.98	0.35	1.72	1.39	0.42	2.31	1.77	0.37	1.92	1.53
White	1837	16.87	0.72	5.67	3.00	0.22	1.19	1.05	0.42	1.86	1.42	0.43	2.43	1.84	0.42	2.07	1.58
Black or African American	4079	13.87	0.66	4.86	2.85	0.21	1.09	0.97	0.29	1.61	1.35	0.40	2.25	1.74	0.35	1.89	1.52
Asian	340	15.18	0.80	6.50	2.93	0.30	1.25	1.05	0.56	2.15	1.42	0.55	2.62	1.77	0.49	2.14	1.52
Native Hawaiian or Other Pacific Islander	22	16.16	0.88	8.55	2.97	0.62	1.67	1.02	0.56	2.20	1.45	0.70	3.39	1.87	0.69	2.60	1.45
Hispanic or Latino	4441	14.13	0.67	4.99	2.88	0.18	1.10	0.99	0.32	1.67	1.37	0.41	2.30	1.76	0.34	1.89	1.53
American Indian or Alaska Native	11	14.41	0.58	4.39	2.86	0.25	1.21	1.04	-0.09	1.43	1.50	0.34	2.17	1.77	0.13	1.43	1.34
Economically Disadvantaged	6654	14.06	0.66	4.89	2.87	0.18	1.08	0.98	0.32	1.66	1.36	0.40	2.26	1.75	0.34	1.89	1.53
Non-Economically Disadvantaged	4394	15.32	0.74	5.81	2.94	0.28	1.21	1.03	0.41	1.82	1.40	0.47	2.49	1.81	0.44	2.07	1.55

8.3 Reliability of Performance Classifications

Several measures of reliability are presented below in Table 8.3.1. KR-20 is an internal consistency measure of the multiple-choice items only.

Table 8.3.1: NJ HSPA KR-20

Test	KR-20
LAL	0.79
Math	0.74

The decision consistency measure is an estimate of how reliably the test classifies students into the performance categories (Partially Proficient, Proficient, and Advanced Proficient) from Livingston and Lewis (1995) is presented in Table 8.3.2 for Language Arts and Table 8.3.3 for Mathematics.

Table 8.3.2: NJ HSPA Language Arts Literacy Decision Consistency

LAL		Observed Score			Observed Total
		Partially Proficient (0 – 19.0)	Proficient (19.5 – 35.0)	Advanced Proficient (35.5 – 54.0)	
True Score	Placement Score				
	Partially Proficient (0 – 19.0)	0.63	0.04	0.00	0.67
	Proficient (20.5 – 35.5)	0.08	0.23	0.01	0.32
	Advanced Proficient (36.0 – 54)	0.00	0.00	0.01	0.01
Expected Total		0.71	0.28	0.01	1.00

Table 8.3.3: NJ HSPA Mathematics Decision Consistency

Math		Observed Score			Observed Total
		Placement Score	Partially Proficient (0 – 19.0)	Proficient (19.5 – 36.0)	
True Score	Partially Proficient (0 – 19.0)	0.81	0.07	0.00	0.88
	Proficient (19.5 – 36.0)	0.02	0.10	0.00	0.12
	Advanced Proficient (36.5 – 48.0)	0.00	0.00	0.00	0.00
	Expected Total	0.83	0.17	0.00	1.00

8.4 Conditional Estimate of Error at Each Cut Score

Table 8.4.1 reports the standard errors of measure (SEs) from WINSTEPS. WINTEPS calculates the standard error at each score point using item response theory and the information function. The equation for the standard error at each value of theta (ability) is given by

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}}$$

where $I(\theta)$ is the information function for a test at θ . For the Rasch model, the information provided by a test at θ is the sum of the item information functions at θ . Plots of the standard errors for all levels of θ may be found in Appendix D.

Table 8.4.1: NJ HSPA Conditional Estimate of Error at Each Cut-Score

Subject	Proficiency Level	Raw Score Cut	Theta Cut	Theta SE	Approximate SE in Raw Points
LAL	Proficient	19.5	0.256	0.25	1.99
	Advanced Proficient	36.0	2.695	0.33	1.51
Math	Proficient	19.5	0.033	0.22	2.27
	Advanced Proficient	36.5	1.717	0.25	1.97

8.5 Rater Reliability

Handscoring Reliability (including Rater Effects)

The basic processes for training of readers, assignment of student papers to readers, rules for requiring a third reader resolution, and other facets of scoring are generally the same for Mathematics and Language Arts Literacy. Two readers read every student response. They are randomly assigned to student work to eliminate reader effects. Although two responses of one student may be read by the same reader (for example, pairs of open-ended items associated with the same reading passage), it is not the case that all of one student's work is read by the same reader. All Measurement Incorporated readers have earned at least one BA or BS college degree, typically related to education, liberal arts, or the specific content area itself. Readers are trained using real student papers. Their initial training involves careful supervision of assigned scores. During training, readers who do not consistently assign the appropriate score are sent for additional training or dismissed, depending on the degree to which they are in error. Because Measurement Incorporated has had a longstanding relationship with NJ DOE through ASK, NJBCT, and HSPA, we have access to a large number of readers who are experienced with these specific assessments, and those are the readers typically chosen for this project. Where two readers' scores differ by more than 1 point (called non-adjacent scores), a third reader, typically a team leader or scoring director makes a third judgment about the student work. As data here show, this is seldom necessary; however, there are specific rules for how these scores should be combined. Appendix C lists these rules.

Although these general procedures are the same for both content areas, the two content areas do differ. The primary differences between the two content areas concern the item types and degree of specificity of the rubrics used for scoring. For Mathematics, all item types are the same. Each Math operational test has 6 open-ended items worth a possible 3 points each. Each item is read by two readers and, where those scores are the same or adjacent, the mean is taken as the final score. When the first two randomly assigned readers differ by more than one point, a third reader resolves the difference. For Language Arts Literacy, there are 4 open-ended items, a pair associated with each of two reading passages. Each of these items is worth a possible 4 points. Third readers resolve non-adjacent scores in a fashion similar to Mathematics. Language Arts Literacy includes two other items types, both of which are writing prompts. A Persuasive Writing Prompt is scored on a scale for a possible 12 points, the sum of two readers' scores, each of which is on a scale with a possible 6 points. An Expository Prompt is scored on a scale for a possible 6 points, the average of two readers' scores, each of which is on a scale with a possible 6 points. All of the Language Arts items, open-ended associated with passages and the Writing prompts, are scored using a holistic rubric that is well-publicized throughout the state as the model for student writing.

For Spring 2015, the following data were obtained. The percent agreement between Reader 1 (R1) and Reader 2 (R2) is very high, as one would expect with a random distribution of forms to readers. Furthermore, the percent of student responses that were scored either with both perfect agreement or adjacent agreement was extremely high, ranging from 99.3% to 99.8%.

Table 8.5.1: NJ HSPA Spring 2015 Reader Agreement Statistics

TEST	Code*	Points	% Perfect	% Adjacent	% Perfect & Adjacent
LAL	OE1	4	91.5	8.2	99.7
LAL	OE2	4	91.9	7.6	99.5
LAL	OE3	4	91.3	8.4	99.7
LAL	OE4	4	89.4	10.0	99.4
LAL	WR1	6	90.6	9.00	99.6
LAL	WR2	12	89.5	9.8	99.3
MATH	OE1	3	91.4	8.4	99.8
MATH	OE2	3	93.8	5.6	99.4
MATH	OE3	3	95.5	3.8	99.3
MATH	OE4	3	89.7	10.0	99.7
MATH	OE5	3	96.1	3.6	99.7
MATH	OE6	3	95.1	4.4	99.5

- Each reader assigns a score out of a possible 6 points and then scores are summed. Reader scores here are, by necessity, on the 6-point scale.
- * Not the actual UIN of an item

Part 9: Validity

The New Jersey Department of Education is developing a comprehensive set of assessments that measure student achievement of the Core Curriculum Content Standards. The validity of the NJ HSPA scores is based on the alignment of the NJ HSPA assessment to the Core Curriculum Content Standards and the knowledge and skills expected of high school students.

The Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999, p.11-12) note the following possible sources of validity evidence:

- Evidence based on test content presented in part 2.1. Construct Validity describes how the test content relates to the construct. Part 2.1 describes the test content that describes the construct.
- Evidence based on internal structure of the test presented in parts 2.2 and 2.3. Content Validity describes the item review and development process. The internal structure of the test is described.
- Evidence based on relations to other variables presented in part 5.3. Concurrent Predictive validity describes the relationship between the test scores and an external test.
- Evidence based on consequences of testing is presented in part 1.2.

For an assessment like NJ HSPA, one intended to measure student's performance in relation to the Core Curriculum Content Standards, content validity evidence is the most relevant and important source of evidence. The section of this technical report on "Test Development," presents validity evidence based on the test content. A description of the test specification development is followed by the procedures for test item development. Details about item writing as well as task, prompt, and passage selection are included. Part 2 delineates the review work of the New Jersey Assessment Content Committees.

Part 10: Reporting

10.1 Cycle I Reports

Individual Student Report

The Individual Student Report (ISR) is a report of how each individual student did on the HSPA. Two copies of the report are produced for every student tested, one for the student's permanent folder after the results are analyzed, and the other for the student's parent/guardian to be shared in a manner determined by the local district.

The scale scores for Language Arts Literacy and Mathematics are provided as well as cluster (standard) data. The Just Proficient Mean is also provided to give the student some information regarding how students who just barely passed did on each cluster (standard). The Just Proficient Mean is the average or mean cluster raw scores for all students (GE, SE, and LEP) in the state whose scale score is 200, i.e., students who are "just proficient." Only first-time eleventh-grade data are used for calculating the Just Proficient Mean. The data for students who are IEP Exempt from Passing and those who took the Braille, large-print, breach, alternate, and special equated versions are excluded from the calculation of these means.

There are eight Summary of State Performance reports, four for Language Arts Literacy and four for Mathematics. Each of these sets of four consists of one per grade, 11th, 11th retained, 12th, and 12th retained. The reports are produced at the school level and provide aggregated data for a test section. Data are provided for total students, general education students, special education students, IEP exempt from passing students, LEP students, and IEP exempt from taking students.

Summary of State Performance

There are twelve Summary of State Performance reports, four for Language Arts Literacy, four for Science, and four for Mathematics. Each of these sets of four consists of one per grade, 11th, 11th retained, 12th, and 12th retained. The reports are produced at the school level and provide aggregated data for a test section. Data are provided for total students, general education students, special education students, IEP exempt from passing students, LEP students, and IEP exempt from taking students.

The report provides the percent and number of students in each proficiency level for each of the groups mentioned above.

New Jersey High School Proficiency Assessment Individual Student Report

Test Date: MARCH 2005
Report Printed: 2/1/2005

County: 88 WOOD
District: 7777 GOOD TOWN
School: 666 PROFICIENCY HIGH

Date of Birth: 01/11/86
Sex: F
Grade: 11
Out of District Placement:

LEP:
SE:
IEP Exempt From Passing:
Title 1:

Answer Folder No: 47987
District/School ID No: 047031
Retest:
Special Form:

Student Name: CARRASQUILLO, REBECCA L

HSPA ID No: 1380289434

Content Area	Your Scale Score	Proficiency Level	
Mathematics	266	ADVANCED PROFICIENT	Partially Proficient / Not Pass: Score BELOW 200
Language Arts Literacy	275	ADVANCED PROFICIENT	Proficient / Pass: Score AT OR ABOVE 200 but BELOW 250
			Advanced Proficient / Pass: Score AT OR ABOVE 250

Mathematics			Language Arts Literacy		
The Mathematics section assesses a student's abilities in the following clusters.			The Language Arts Literacy section assesses a student's abilities in the following clusters.		
Cluster	Your Raw Score	Just Proficient Means	Cluster	Your Raw Score	Just Proficient Means
Number & Numerical Operations	6.0 out of 7	9.9	Writing	16.0 out of 18	9.9
Geometry & Measurement	10.0 out of 12	9.9	Reading	28.0 out of 36	9.9
Patterns & Algebra	10.0 out of 15	9.9	-----		
Data Analysis, Probability & Discrete Mathematics	13.0 out of 14	9.9	Interpreting Text	8.0 out of 11	9.9
-----			Analyzing / Critiquing Text	20.0 out of 25	9.9
Knowledge	39.0 out of 48	9.9			
Mathematical Processes - Problem Solving	24.5 out of 26	9.9			

ABOUT THE HIGH SCHOOL PROFICIENCY ASSESSMENT (HSPA)

The High School Proficiency Assessment (HSPA) was administered over three days. The HSPA measures student knowledge and skills in two areas: Language Arts Literacy and Mathematics. Students who entered eleventh grade for the first time on or after September 1, 2001, must pass both sections of the HSPA as a graduation requirement. The HSPA was designed with the involvement of educators, business representatives, and parents to ensure that your child has acquired the types of skills and knowledge that will help prepare him or her for the future -- whether that includes college, employment, or the military.

The **Language Arts Literacy** section measures skills in two areas: reading and writing. The reading component requires students to read passages and answer related questions about each passage. Most of the questions are multiple-choice; however, some questions require students to provide written responses using their own words. These are referred to as "open-ended" questions and were scored by highly trained scorers. Reading passages test comprehension, both literal and inferential. Literal comprehension is the ability to understand the actual meaning of written words. Inferential comprehension is the ability to use careful reasoning to extend understanding of the communication beyond the literal meaning of the words themselves.

The writing component requires students to respond to two writing prompts. One prompt presents a photograph and requires students to create a story based on features or elements of the photo. The other prompt provides a topic and requires students to write a persuasive essay based on that topic. These two tasks measure students' ability to construct meaning in sustained written responses. Each of your child's responses was read by two readers trained to score student writing using criterion-based holistic scoring. This scoring method requires readers to examine the writing for the following characteristics: content and organization, construction of sentences, use of the English Language, and writing mechanics. Each of two readers gives the writing a score that ranges from 1 (the lowest) to 6 (the highest). Low points indicate problems with written language, and high points indicate satisfactory performance.

The **Mathematics** section tests students' knowledge of the following skills: number and numerical operations; geometry and measurement; patterns and algebra; and data analysis, probability, and discrete mathematics. Most of the Mathematics section consists of multiple-choice questions. Some questions are open-ended and were scored by highly trained scorers.

HOW TO READ THIS REPORT

The **Individual Student Report** presents your child's Language Arts Literacy and Mathematics scores on the HSPA. This report is available only to parents, guardians, students, and authorized school officials. If you have any questions about this report, you should contact your child's guidance counselor or principal.

Your child's name, birthdate, and other personal information are at the top of the report. An asterisk (*) may appear in the name, birthdate, or district/school ID number if the circles for more than one letter or number were filled in the same column on the answer folder.

The total HSPA Language Arts Literacy and Mathematics scores are reported as scale scores with a range from 100 to 300. The HSPA passing score is 200. To meet the state's graduation testing requirement for a high school diploma, a student must score at or above the passing score for each section of the HSPA. If your child does not pass one or more sections, he or she will be provided with remediation, tested again, and given a Special Review Assessment in the twelfth grade. Students are not retested in sections they have already passed.

Your child's total HSPA Language Arts Literacy and HSPA Mathematics scores are presented in the box on the top half of the report. Your child's scale score for each HSPA section is printed in the column labeled **Your Scale Score**. To the right of the scale score is a column labeled **Proficiency Level**. If the scale score is below 200, your child is "Partially Proficient" in that content area. If the scale score is 200 to 249, your child is "Proficient" in that content area. If the scale score is 250 to 300, your child is "Advanced Proficient" in that content area. Scores below 200 indicate a need for additional instructional assistance. However, as with any single test score, HSPA results should not be used as the sole basis for instructional decisions.

Additional information to assist in identifying your child's strengths and weaknesses is presented at the bottom half of the report. Cluster-level results show how your child performed on the items that measure particular knowledge and skills. Although an item on the HSPA can contribute to more than one cluster (for example, reading and interpreting text), each item is counted only once to calculate the scale score.

For each cluster, the column labeled **Your Raw Score** presents the number of points your child achieved. The column labeled **Just Proficient Means** is a yardstick against which you can measure your child's performance for each cluster. Each **Just Proficient Mean** is the average raw score for all students in the state whose scale score is 200 for the particular content area. If your child scored at or above the **Just Proficient Mean**, this cluster is an area of possible strength for your child. If your child scored below the **Just Proficient Mean**, your child is likely to need additional help in this cluster.

A notation may appear if, for some reason, your child's answer folder was not scored. No data will appear under **Your Raw Score** and **Your Scale Score**. Instead, the report will indicate one of the following: Not Present, IEP Exempt From Taking, Not Scored, or Void. A void may be assigned for one of the following reasons:

- Void 1. Your child became ill during testing.
- Void 2. Your child refused to test, cheated, or was disruptive during testing.
- Void 3. Other reason:
 - a) Your child tested but had already passed a HSPA section;
 - b) Your child tested but was not in the correct grade; or
 - c) Your child tested a section(s) twice.
- Void 4. Your child attempted an insufficient number of items.

10.2 Cycle II Reports

Cluster Means for Students with Valid Scale Scores Report

The Cluster Means for Students with Valid Scale Scores Report are produced for each content area, Mathematics and Language Arts Literacy. It provides the means for each subgroup of students (GE, SE, LEP, IEP Exempt from Passing, and Total) for each of the content area clusters. Mean performance is provided for the school, district, and state, as well as for the statewide DFG representing the district. If the district is a Special Needs district, cluster means are shown for the statewide Special Needs population and the statewide non-Special Needs population. Cluster performance for students who took the Braille, large-print, breach, alternate, and special equated versions of the assessment are excluded from the report.

Performance by Demographic Group Report

The Performance by Demographic Group Report is a one-page report that presents test results for both content sections of the HSPA. The test results are broken down within the report by student subgroup (GE, SE, LEP, IEP Exempt from Passing, and Total), gender, migrant status, ethnicity, and economic status. Individual reports are produced for the school, district, statewide population, and the DFG that includes your district. For Special Needs districts, reports are also produced for the statewide Special Needs population, and the statewide non-Special Needs population.

Special School Report

The Special School Report is produced at the school level only. The use of a special code on the HSPA answer folder allows the district to select achievement data on particular groups of students. There are separate reports for Mathematics and Language Arts Literacy, and each provides both performance data and cluster means. The performance data section of the report is broken down by student subgroup (GE, SE, LEP, IEP Exempt from Passing, and Total) while the cluster section provides the means for the selected students on each of the clusters.

10.3 State Summary Report (Reporting Rules)

Reporting Rules for 2015 Electronic File of State Summary with Suppression

The reporting rules used for 2015 are the same as those finalized by the New Jersey Department of Education (DOE) in 2009 for the State Summary Reports for HSPA and NJ ASK. The rules are listed below and aim at protecting confidential information when reports are released to the public. Each contractor needs to adopt the rules when generating suppressed state summary report files.

A. THE FOLLOWING RULES APPLY TO ALL UNSUPPRESSED RECORDS:

1. (State Summary Only) When the enrollment of a group is 0, report this number, and leave all other fields for that group blank.

2. (State Summary Only) When the enrollment of a group is not 0 and the number of valid scores is 0, report all n-counts, and leave all fields pertaining to proficiency range and mean scale score blank for that group.
3. (Report Card Only) When the number of valid scores for a group = 0, report all n-counts, and leave all fields pertaining to proficiency range and mean scale scores blank for that group.

B. THE FOLLOWING RULES APPLY TO ALL SUPPRESSED RECORDS:

1. When the results of a group are suppressed, replace all the data for that group with single asterisks.
2. When the number of students with valid scale scores in a reporting group is greater than 0 and less than 11, suppress all data for that group.
3. When the percentage of partially proficient scores in a reporting group is greater than 90% (90.05% or greater), suppress all data for that group.

C. THE FOLLOWING RULES APPLY TO SPECIFIC CASES: (school and district level)

1. General Education, SE, LEP (Report Card only)
 - a. When one or two of the three groups has greater than 10 valid scores and only one of the three groups has greater than 0 and less than 3 valid scores, suppress all data for SE and LEP.
 - b. When only one of these three groups has greater than 90% partially proficient scores, suppress all data for SE and LEP.
2. General Education, SE, Current LEP (State Summary only)
 - a. When one or two of the three groups has greater than 10 valid scores and only one of the three groups has greater than 0 and less than 3 valid scores, suppress all data for SE, current LEP and the sum of current plus former LEP.
 - b. When only one of these three groups has greater than 90% partially proficient scores, suppress all data for SE, current LEP and the sum of current plus former LEP.
3. Ethnicity
 - a. When any racial/ethnic group has greater than 10 valid scores and only one other group has greater than 0 and less than 3 valid scores, suppress all data for Native American, Pacific Islander, and "Other."
 - b. When any racial/ethnic group has greater than 90% partially proficient scores, suppress all data for Native American, Pacific Islander, and "Other."

4. Gender

- a. When one gender has greater than 0 and less than 3 valid scores, suppress all data for the other gender.
- b. When one gender has greater than 90% partially proficient scores, suppress all data for the other gender.

5. LEP status (State Summary only)

- a. When the current LEP group has greater than 10 scores and the former LEP group has greater than 0 and less than 3 valid scores, suppress all data for the sum of current plus former LEP.
- b. When the former LEP group has greater than 90% partially proficient scores, suppress all data for the sum of current plus former LEP
- c. When the former LEP group has greater than 10 scores and the current LEP group has greater than 0 and less than 3 valid scores, suppress all data for the sum of current plus former LEP.
- d. When the current LEP group has greater than 90% partially proficient scores, suppress all data for the sum of current plus former LEP

6. Migrant Status (State Summary only)

- a. When the migrant group has greater than 0 and less than 3 valid scores, suppress all data for the non-migrant group.
- b. When the migrant group has greater than 90% partially proficient scores, suppress all data for the non-migrant group.

7. IEP exempt from passing and Non-IEP exempt from passing (HSPA State Summary only)

- a. When one of the above two groups has greater than 0 and less than 3 valid scores, suppress all data for the other group.
- b. When one of the above two groups has greater than 90% partially proficient scores, suppress all data for the other group.

Record Inclusion Rules

The following records are included in the state summary data files:

- Statewide
- Statewide Non-Special Needs
- Statewide Special Needs
- 1 record per DFG, excluding DFGs O and S (see below)
- 1 record per district with participating students, except in DFGs O and S
- 1 record per school with participating students, except in DFGs O and S

DFG O includes DOC, DHS, DCF, JJC and private schools. R represents Charter schools. S represents educational services commissions and special services districts.

DFG N (internal code, “Not Designated”) is the code used for districts that lost a designated DFG. The present N designation includes Lakewood and Deal districts. V represents vocational schools. The record for DFG N just contains the DFG code; it does not contain any data.

All “IEP Exempt From Taking” columns will be blank except for the enrollment counts.

The percent proficient and scale score mean columns will be blank in the “SE-Accommodations” section.

All science columns (including Number Enrolled) will be blank.

Record Sort Order

The first three records are as follows, in the order specified:

1. Statewide
2. Non-Special Needs
3. Special Needs

These records are followed by the first DFG record when ordered alphabetically, then its districts in alphabetical order by name, with each district record followed by its school records in alphabetical order. Each subsequent DFG, and its districts and schools, follow after that. The county name is not used in alphabetizing the district or school records.

4. DFG A
5. Monmouth County - Asbury Park District
6. Monmouth County - Asbury Park District - Asbury Park School
7. Atlantic County - Atlantic City District
8. Atlantic County - Atlantic City District - Atlantic City School
9. Cumberland County - Bridgeton District
10. Cumberland County - Bridgeton District - Bridgeton School
11. etc.
12. DFG B

13. Hudson County - Bayonne District
14. Hudson County - Bayonne District - Bayonne School
15. Burlington County - Burlington City District
16. Burlington County - Burlington City District - Burlington City School
17. Middlesex - Carteret District
18. Middlesex - Carteret District - Carteret School
19. etc.

Excel File Features

The Excel file version of the state summary data file uses the Excel template provided by the New Jersey Department of Education in 2005, which is formatted for convenient printing. Note that CDS codes should be treated as character values in the Excel file so that leading zeroes are shown. The first row of the template includes descriptive column headers. Other than that, the same records appear in both Excel and text versions, row for row. The numbers are exactly the same between the text and Excel files except for two formatting differences:

1. Implied decimal points are used in the text file. Actual decimal points are used in the Excel version.
2. In the text file, numbers are shown with enough leading zeroes to fill the entire column. 85.4% should be 0854 and 1.2% should be 0012. In the Excel version, there are no leading zeroes, except for one leading zero for numbers less than 1%. 85.4% should be 85.4 and 0.2% should be 0.2.

The Excel file version of the state summary data file should be provided in one workbook containing five worksheets, which follow the same order as the file layout.

The summary data should be organized in worksheets with the corresponding worksheet names as follows:

- Total and Instructional Groups
- Gender
- Migrant
- Ethnic
- Economic

The instructional Groups for HSPA are GE, SE, SE-Accommodations, LEP (Current + Former), LEP Current, LEP Former, IEP Exempt From Taking, IEP Exempt From Passing, and IEP Not Exempt From Passing.

The first eight columns (“Co”, “Dist”, “Sch”, “Cnty Name”, “Dist Name”, “School Name”, “DFG”, “Spec Needs”), which uniquely identify the record, will be repeated on all worksheets.

Freeze the first row and first six columns so that they remain visible as the user scrolls across or down the worksheet.

10.4 Interpreting Reports

Raw Score Data for School, District, and State – Data include the mean number of points obtained by all the listed student groups (Total Students, GE, SE, LEP, IEP Exempt from Passing, and Title I) for each cluster in the school, district, and state. HSPA answer folders coded as void are excluded from these means. The data for students who took the Braille, large-print, breach, alternate, and special equated versions are also excluded.

Raw Score Data for Special Needs and Non-Special Needs – Data include the mean number of points obtained by all the listed student groups (Total Students, GE, SE, LEP, IEP Exempt from Passing, and Title I) for each cluster in the 31 Special Needs or Abbott districts and the mean number of points for these same listed student groups for each cluster in all other non-Special Needs districts.

Raw Score Data for DFG – Data include the mean number of points obtained by all the listed student groups (Total Students, GE, SE, LEP, IEP Exempt from Passing, and Title I) for each cluster in the same DFG as the district listed on this report.

Scale Scores:

The total HSPA Mathematics and Language Arts Literacy scores are reported as scale scores with a range of 100 to 300. Please note that 100 and 300 are a theoretical floor and ceiling and may not be actually observed. The scale score of 250 is the cut point between Proficient students and Advanced Proficient students. The scale score of 200 is the cut point between Proficient students and Partially Proficient students. The score ranges are as follows:

Advanced Proficient/Pass 250–300

Proficient/Pass 200–249

Partially Proficient/Not Pass 100–199

The scores of students who are included in the Partially Proficient level are considered to be below the state minimum level of proficiency. These students may need additional instructional support, which could be in the form of individual or programmatic intervention. It is important that districts consider multiple measures with all students before making decisions about students' instructional placements.

10.5 Quality Control in Reporting

Quality control procedures for score reporting

MI fully recognizes the importance of error free score reporting, and employs stringent quality control procedures to ensure that reporting on all levels is completely accurate. MI thoroughly tests, reviews, and proofreads all reporting deliverables prior to any delivery to NJ DOE or other parties such as other state offices and districts.

MI uses structured testing methodologies to test all software programs that are used in the production of score reports. All scanning and scoring programs are fully tested and reviewed prior to the start of live material processing and are continually monitored throughout the process. MI's

QA staff develops and utilizes independent queries to validate all software programs and programmatically produced deliverables for reporting. Before we begin to report any results, we know that the data are accurate as a first step. Each program is tested to ensure that all data are included or excluded as appropriate (with particular attention to Braille and Large Print students as well as any other special equating situations) and to ensure that all programmatic calculations are performed accurately and according to the reporting rules provided by the State Contract Manager. We verify that all formatting rules are followed and that reports are presented as they were designed by MI and NJ DOE with all appropriate fonts, footnotes, line separations, sections, and headings. We test all aspects of the reports to ensure that valid values are verified, all valid codes are included on all student records, correct scores are reflected and are attributed to the correct student, cluster scores are accurately aggregated and totaled, and appropriate student totals are reported in all aggregate reports. In addition, during phase II of Cycle I Quality Control (QC) review, we ensure that all appropriate record changes made by districts have been applied. The testing process described above is inclusive of data files, electronic reports and printed reports.

MI works cooperatively with NJ DOE to arrange for NJ DOE QC review of selected districts, schools, and students prior to release of any score reporting. Currently with HSPA, we assist the NJ DOE staff with this QC review in various ways. We developed the guidelines for QC sample selection in conjunction with NJ DOE staff. We pre-select and propose a QC sample based on these guidelines. After the initial proposed selection is made, the list is sent to NJ DOE for approval and any additions that they wish to make. The QC sample consists of students enrolled in 8-9 QC districts plus miscellaneous individual student records. The QC districts always include the following:

- A Charter school,
- At least one school district with multiple schools, and
- Districts taking the alternate form and any other districts with students receiving any special equating due to testing irregularities.

Altogether, the QC districts are distributed over the full range of DFGs, but do not necessarily represent each DFG. In March administrations for HSPA when there are multiple types of answer folders and the layout of the multiple-choice grids is different, we select at least one district with each layout. A different set of districts is chosen for each administration.

In addition to the QC districts, miscellaneous individual student records also include:

- A minimum of five Braille and five large-print students testing in all content areas since these students have certain items omitted, leading to different maximum cluster scores. If there are not enough Braille and large print students in the QC districts, more students from other districts are added.
- Any miscellaneous problem answer folders that NJ DOE wants to monitor. These may be added before or after the initial QC sample has been selected.

Since we constantly review our QC process after each test administration, we have been able to streamline the process and make it much more efficient and effective than it was when we began work on the HSPA contract in 2002. We have found that the method we currently use to select the sample for HSPA provides the best coverage of test scenarios for review and allows us to identify

easily and efficiently students who represent all scenarios. It also allows the NJ DOE staff to review fewer students to determine the correctness of our processing and programs. Through our improvements to the QC process, we have realized a significant decrease in the amount of time needed to complete the review. We hope that this has alleviated the strain on NJ DOE resources that currently exists.

During the QC process, live answer folders are hand-scored for multiple-choice (MC) items using scoring masks, and reader score sheets are reviewed and compared to student records to ensure that the hand scores are appropriately applied. A selection of students is presented to ensure coverage of each type of demographic coding scenario as well as any overrides that are done by MI according to coding rules as developed by NJ DOE.

Live reports are reviewed for the selection of students, schools, and districts. Reports are available to the State Contract Manager immediately following review and approval by MI QA staff. MI provides printed and electronic copies of a sample of all reports to the State Contract Manager for each phase of NJ DOE QC review. These reports are made available via download from the secure FTP site. All reports are provided in a timely manner for each phase and in advance whenever possible.

MI programming and QA staff are readily available during all NJ DOE QC reviews to assist however necessary. At the end of the day, our goal is for NJ DOE and MI to feel confident in knowing that each and every report produced and distributed is accurate and complete. Accuracy and timeliness of all reporting are the primary concern of MI staff.

Preliminary reporting quality control sample.

MI works with NJ DOE to select certain districts and other special educational entities to include in the quality control sample for score reporting. As discussed in the previous section, this sample represents regular schools as well as Charter schools, educational programs administered by the state, and private schools for the disabled. The sample at each grade level has up to 10 schools in eight regular districts and up to three special districts or state facilities.

MI makes arrangements for the early processing of answer folders after the first spring week of testing so that they can be scanned and scored first. This gives us additional time to prepare the reports for examination by our QA staff prior to NJ DOE review. This also allowed us to introduce a limited amount of live data to our systems, so that we can troubleshoot any problematic situations that become apparent prior to operating at higher capacity. Once we scan these documents, we save a copy of the data, which is used to create the sample reports. Using a copy allows other processes such as record change requests to be applied to the original data without causing changes to the quality control reports.

After MI's QA staff checks the reports for accuracy, a complete set of Cycle I reports for these students is generated for comparison to the actual answer documents. MI provides original answer folders and photocopies of the answer documents and scannable reader score sheets so that NJ DOE staff can compare the gridding of demographic and multiple choice bubbles, and the open-ended and writing task scores to the score reports.

The QC review for the March 2009 HSPA administration went very well and all data and reports were approved for release at the conclusion of the visit. We then began to print reports for on-time distribution to districts.

References

Linacre, J. M. (2004). Winsteps® Rasch measurement computer program. Beaverton, Oregon: Winsteps.com

Livingston, S.A., & Lewis, C. (1995). Estimating the consistency and accuracy of classifications based on test scores. *Journal of Educational Measurement*, 32, 179-197

Appendix A: Test Specifications from the February 1998 Report

Test Specifications from the February 1998 Report

Language Arts Literacy Test Keywords and Definitions

Working with Text (ESPA) – Interpreting Text (GEPA and HSPA)

Working with Text refers to those activities in which students use strategies to interpret or reformulate meaning from the text. Questions and tasks with this focus will ask students to identify main ideas, supporting details, directions, paraphrasing, text organization, and purposes for reading, listening, or viewing.

Analyzing/Critiquing Text

Analyzing/Critiquing Text refers to those activities in which students use strategies to analyze and critique the text. Students will pose or respond to questions that enhance their understanding, predict tentative meanings, and draw conclusions or form opinions about the text and the author's techniques. Questions and tasks that focus on this kind of analysis will ask students to identify or explain the fundamentals and the nuances contributed by textual conventions and literary elements.

Extending Understanding of the Text

Extending Understanding of the Text refers to those activities in which students use text already generated, that is, informational and everyday texts, as a springboard for generating their own work and ideas. It is a self-contained component for which students will read a passage to make decisions, solve a problem, and create original work through a writing project that is designed to extend their understanding of the text. This section of the assessment is constructed to enable students to shape their written response with what they have learned through their reading and from what they know through prior experience.

Working with Text or Interpreting Text

Interpreting text involves activities and strategies that contribute to reformulating meaning, including:

- developing explanations and extrapolating information,
- developing specific purposes and inferring purposes, and
- planning and recognizing the organization of texts.

Analyzing and Critiquing Text

Students will be able to pose or respond to questions in ways that enhance their and others' understandings of the text. They will predict tentative meanings of texts and plan texts as temporary thinking on their way to drawing conclusions or forming opinions. These conclusions and opinions will eventually take on more formal expressions when students move to extending understanding of text. Through this process of analysis and critique, students will understand both the functions and nuances of textual conventions and literary elements.

Extending Understanding of the Text

Students will be able to create original works. Some of these works are textual, more finished products that they can make available to specific audiences and/or for specific purposes. Some extensions of understanding result in the reader appreciating a text or its features, considering other related texts, or interacting with others' related ideas, all of which extend literacy. Some extensions of understanding will lead students to take action. This action will include problem-solving, making decisions, and creating an original work, which may lead to heightened social awareness and action.

Definition of Text

The term *text*, as used in this directory, is consistent with the use of the term in the Language Arts Literacy component of the New Jersey *Core Curriculum Content Standards*. Text refers to any printed or oral use of language. It also includes any visual communication that we "read."

Narrative Text

For the purposes of this assessment, "narrative text" is defined as literature written primarily to tell a story. Good narrative literature, which establishes or develops a conflict, addresses common aspects of human existence. Because appropriate literature may contain unsettling or disturbing issues or events, text selected for the assessment will provide a positive resolution and affirm the dignity of the human spirit. Selections will provide students with opportunities to grow intellectually, socially, and emotionally as they consider universal themes and diverse cultures and perspectives.

Narrative passages will be selected from previously published literature of between 1200 and 1800 words for ESPA and 2100 and 3300 words for GEPA and HSPA. Students will respond to open-ended and multiple-choice questions about those passages. The texts will have a strong thematic focus, follow traditional narrative structure, and contain the following elements:

- significant themes that are age- and grade-level appropriate;
- a clearly identifiable problem/conflict and resolution;
- a well-organized plot with clearly developed major events;
- well-developed characters;
- settings integral to the plot;
- literary devices, such as imagery and foreshadowing; and
- a range of vocabulary for which adequate context is provided.

Persuasive Text (GEPA and HSPA only)

"Persuasive text is defined as text in which the writer attempts to sway the reader to a specific point of view. Because persuasive writing is based on a personal vision, it is inherently controversial. Exploring these controversial issues develops and enhances students' critical thinking skills. Selected texts, which will introduce topics that are consistent with topics presented in other sections of the assessment unit, will address issues that challenge, broaden, and enrich their perspectives.

Persuasive passages will be selected from previously published texts, such as essays, speeches, book and movie reviews, editorials, letters, advertisements, charitable campaign appeals, and political literature. Students will respond to open-ended and multiple-choice questions about those passages. Persuasive text of between 1000 and 1600 words may be excerpted or used in full.

Persuasive text will contain the following elements:

- topics that are age- and grade-level appropriate;
- a clear focus;
- elaboration using facts and opinions;
- persuasive techniques, including but not limited to, propaganda, connotative and figurative language, and rhetorical devices; and
- a range of vocabulary for which adequate context is provided.

Mathematics Keywords and Definitions

Knowledge and Skills

The purpose of New Jersey's statewide assessment program is to measure what students at benchmark grade levels know and are able to do.

Student knowledge can be broken down into conceptual knowledge (including facts learned) and procedural knowledge (including, in mathematics, ability to perform certain algorithmic processes). Problem – solving skills include student ability to select and apply the knowledge learned and algorithmic processes mastered to rich, engaging situations.

The framework portrayed on the preceding page provides a structure for the eighth-grade and eleventh-grade tests. It characterizes the mathematics to be assessed not only as involving either knowledge or problem-solving skills, but also as falling into one or more of four content standards.

Content Standards

New Jersey's eighth-grade mathematics test assess knowledge and skills in four content areas of standards:

- I** Number and Numerical Operations
- II** Geometry and Measurement
- III** Patterns and Algebra
- IV** Data Analysis, Probability, and Discrete Mathematics

These standards unavoidably contain some overlapping content, since mathematical topics are not disconnected but are part of an interconnected whole.

Mathematics Strands, by Standard

I. Number and Numerical Operations

- A. Number Sense
- B. Numerical Operations
- C. Estimation

II. Geometry and Measurement

- A. Geometric Properties
- B. Transforming Shapes
- C. Coordinate Geometry
- D. Units of Measurement
- E. Measuring Geometric Objects

III. Patterns and Algebra

- A. Patterns and Relationships
- B. Functions
- C. Modeling
- D. Procedures

IV. Data Analysis, Probability, and Discrete Mathematics

- A. Data Analysis (Statistics)
- B. Probability
- C. Discrete Mathematics—Systematic Listing and Counting
- D. Discrete Mathematics—Vertex-Edge Graphs and Algorithms

Appendix B: Raw to Scale Score Conversions and Frequency Distributions

Table B.1: Language Arts Literacy Raw to Scale Score Conversion and Frequency Distribution

Raw Score	Theta	SE	SS	Freq	%	Cum. Freq	Cum. %
0.0	-4.85	1.80	100	3	0.08	3	0.08
0.5	-3.69	0.97	100	0	0.00	3	0.08
1.0	-3.04	0.69	100	5	0.14	8	0.22
1.5	-2.66	0.56	100	1	0.03	9	0.25
2.0	-2.38	0.50	102	13	0.36	22	0.61
2.5	-2.15	0.45	107	3	0.08	25	0.69
3.0	-1.97	0.42	112	20	0.56	45	1.25
3.5	-1.80	0.39	117	5	0.14	50	1.39
4.0	-1.66	0.37	123	35	0.97	85	2.36
4.5	-1.53	0.34	128	7	0.19	92	2.56
5.0	-1.42	0.32	133	43	1.19	135	3.75
5.5	-1.33	0.30	138	19	0.53	154	4.28
6.0	-1.25	0.28	142	45	1.25	199	5.53
6.5	-1.17	0.26	146	27	0.75	226	6.28
7.0	-1.11	0.25	149	55	1.53	281	7.81
7.5	-1.05	0.24	152	26	0.72	307	8.53
8.0	-0.99	0.23	154	64	1.78	371	10.31
8.5	-0.94	0.22	155	46	1.28	417	11.59
9.0	-0.89	0.22	158	70	1.94	487	13.53
9.5	-0.84	0.22	160	47	1.31	534	14.84
10.0	-0.80	0.21	162	89	2.47	623	17.31
10.5	-0.75	0.21	164	60	1.67	683	18.98
11.0	-0.71	0.21	166	89	2.47	772	21.45
11.5	-0.66	0.21	167	77	2.14	849	23.59
12.0	-0.62	0.22	169	94	2.61	943	26.20
12.5	-0.57	0.22	171	87	2.42	1030	28.62
13.0	-0.52	0.22	173	120	3.33	1150	31.95
13.5	-0.47	0.23	175	116	3.22	1266	35.18
14.0	-0.42	0.23	177	116	3.22	1382	38.40
14.5	-0.36	0.24	179	113	3.14	1495	41.54
15.0	-0.30	0.24	181	132	3.67	1627	45.21
15.5	-0.25	0.24	183	114	3.17	1741	48.37
16.0	-0.19	0.25	185	118	3.28	1859	51.65
16.5	-0.12	0.25	187	112	3.11	1971	54.77
17.0	-0.06	0.25	189	145	4.03	2116	58.79
17.5	0.00	0.25	191	115	3.20	2231	61.99
18.0	0.07	0.25	193	110	3.06	2341	65.05
18.5	0.13	0.25	195	111	3.08	2452	68.13
19.0	0.19	0.25	197	92	2.56	2544	70.69
19.5	0.26	0.25	200	99	2.75	2643	73.44
20.0	0.32	0.25	201	77	2.14	2720	75.58
20.5	0.38	0.25	203	79	2.20	2799	77.77
21.0	0.45	0.25	205	71	1.97	2870	79.74
21.5	0.51	0.25	207	62	1.72	2932	81.47
22.0	0.57	0.25	209	51	1.42	2983	82.88
22.5	0.64	0.25	210	61	1.69	3044	84.58
23.0	0.70	0.25	212	45	1.25	3089	85.83
23.5	0.76	0.25	214	40	1.11	3129	86.94
24.0	0.83	0.25	215	49	1.36	3178	88.30
24.5	0.89	0.25	217	29	0.81	3207	89.11
25.0	0.95	0.25	218	35	0.97	3242	90.08
25.5	1.02	0.25	220	30	0.83	3272	90.91
26.0	1.08	0.26	222	22	0.61	3294	91.53

Raw Score	Theta	SE	SS	Freq	%	Cum. Freq	Cum. %
26.5	1.15	0.26	223	29	0.81	3323	92.33
27.0	1.21	0.26	225	24	0.67	3347	93.00
27.5	1.28	0.26	226	23	0.64	3370	93.64
28.0	1.35	0.26	227	19	0.53	3389	94.17
28.5	1.42	0.26	229	14	0.39	3403	94.55
29.0	1.49	0.27	230	16	0.44	3419	95.00
29.5	1.56	0.27	232	16	0.44	3435	95.44
30.0	1.63	0.27	233	14	0.39	3449	95.83
30.5	1.70	0.27	234	11	0.31	3460	96.14
31.0	1.78	0.28	236	16	0.44	3476	96.58
31.5	1.86	0.28	237	7	0.19	3483	96.78
32.0	1.94	0.29	238	18	0.50	3501	97.28
32.5	2.02	0.29	240	11	0.31	3512	97.58
33.0	2.11	0.30	241	9	0.25	3521	97.83
33.5	2.20	0.30	243	11	0.31	3532	98.14
34.0	2.29	0.31	244	6	0.17	3538	98.31
34.5	2.38	0.31	245	4	0.11	3542	98.42
35.0	2.48	0.32	247	6	0.17	3548	98.58
35.5	2.59	0.33	248	7	0.19	3555	98.78
36.0	2.70	0.33	250	9	0.25	3564	99.03
36.5	2.81	0.34	252	3	0.08	3567	99.11
37.0	2.93	0.35	253	6	0.17	3573	99.28
37.5	3.05	0.36	255	4	0.11	3577	99.39
38.0	3.18	0.36	256	3	0.08	3580	99.47
38.5	3.31	0.37	258	1	0.03	3581	99.50
39.0	3.45	0.37	259	0	0.00	3581	99.50
39.5	3.59	0.38	261	4	0.11	3585	99.61
40.0	3.74	0.38	263	2	0.06	3587	99.67
40.5	3.89	0.39	264	2	0.06	3589	99.72
41.0	4.04	0.39	266	3	0.08	3592	99.81
41.5	4.19	0.39	267	1	0.03	3593	99.83
42.0	4.35	0.40	269	1	0.03	3594	99.86
42.5	4.51	0.40	270	1	0.03	3595	99.89
43.0	4.67	0.41	272	1	0.03	3596	99.92
43.5	4.84	0.41	273	2	0.06	3598	99.97
44.0	5.01	0.41	275	0	0.00	3598	99.97
44.5	5.18	0.41	276	0	0.00	3598	99.97
45.0	5.35	0.42	277	0	0.00	3598	99.97
45.5	5.52	0.42	279	0	0.00	3598	99.97
46.0	5.70	0.42	280	0	0.00	3598	99.97
46.5	5.87	0.42	282	0	0.00	3598	99.97
47.0	6.05	0.42	283	0	0.00	3598	99.97
47.5	6.23	0.43	284	0	0.00	3598	99.97
48.0	6.42	0.43	285	0	0.00	3598	99.97
48.5	6.61	0.44	287	0	0.00	3598	99.97
49.0	6.80	0.45	288	0	0.00	3598	99.97
49.5	7.02	0.47	289	0	0.00	3598	99.97
50.0	7.24	0.49	290	0	0.00	3598	99.97
50.5	7.50	0.52	291	0	0.00	3598	99.97
51.0	7.79	0.55	293	1	0.03	3599	100.00
51.5	8.12	0.60	294	0	0.00	3599	100.00
52.0	8.50	0.65	295	0	0.00	3599	100.00
52.5	8.96	0.71	296	0	0.00	3599	100.00
53.0	9.53	0.81	298	0	0.00	3599	100.00
53.5	10.38	1.07	300	0	0.00	3599	100.00
54.0	11.69	1.87	300	0	0.00	3599	100.00

Table B.2: Mathematics Raw to Scale Score Conversion and Frequency Distribution

Raw Score	Theta	SE	SS	Freq	%	Cum. Freq	Cum. %
0.0	-5.92	1.84	132	0	0.00	0	0.00
0.5	-4.69	1.02	133	0	0.00	0	0.00
1.0	-3.96	0.73	135	2	0.02	2	0.02
1.5	-3.51	0.61	136	0	0.00	2	0.02
2.0	-3.19	0.54	137	6	0.05	8	0.07
2.5	-2.93	0.49	139	0	0.00	8	0.07
3.0	-2.71	0.45	140	14	0.13	22	0.20
3.5	-2.52	0.42	142	1	0.01	23	0.21
4.0	-2.35	0.40	143	60	0.54	83	0.75
4.5	-2.20	0.38	145	3	0.03	86	0.78
5.0	-2.07	0.36	147	106	0.96	192	1.74
5.5	-1.94	0.35	148	11	0.10	203	1.84
6.0	-1.82	0.34	150	195	1.77	398	3.60
6.5	-1.71	0.33	152	26	0.24	424	3.84
7.0	-1.61	0.32	154	300	2.72	724	6.55
7.5	-1.51	0.31	155	41	0.37	765	6.92
8.0	-1.42	0.30	157	423	3.83	1188	10.75
8.5	-1.33	0.29	159	83	0.75	1271	11.50
9.0	-1.24	0.29	161	501	4.53	1772	16.04
9.5	-1.16	0.28	163	92	0.83	1864	16.87
10.0	-1.09	0.28	165	620	5.61	2484	22.48
10.5	-1.01	0.27	166	138	1.25	2622	23.73
11.0	-0.94	0.27	168	685	6.20	3307	29.93
11.5	-0.87	0.26	170	148	1.34	3455	31.27
12.0	-0.80	0.26	172	645	5.84	4100	37.11
12.5	-0.74	0.25	174	163	1.48	4263	38.59
13.0	-0.67	0.25	176	653	5.91	4916	44.50
13.5	-0.61	0.25	178	209	1.89	5125	46.39
14.0	-0.55	0.24	179	618	5.59	5743	51.98
14.5	-0.49	0.24	181	247	2.24	5990	54.22
15.0	-0.43	0.24	183	590	5.34	6580	59.56
15.5	-0.38	0.24	185	225	2.04	6805	61.59
16.0	-0.32	0.23	187	525	4.75	7330	66.35
16.5	-0.27	0.23	189	221	2.00	7551	68.35
17.0	-0.22	0.23	190	462	4.18	8013	72.53
17.5	-0.17	0.23	192	248	2.24	8261	74.77
18.0	-0.12	0.22	194	378	3.42	8639	78.20
18.5	-0.07	0.22	196	133	1.20	8772	79.40
19.0	-0.02	0.22	197	381	3.45	9153	82.85
19.5	0.03	0.22	200	174	1.57	9327	84.42
20.0	0.08	0.22	201	295	2.67	9622	87.09
20.5	0.13	0.22	203	128	1.16	9750	88.25
21.0	0.18	0.22	204	221	2.00	9971	90.25
21.5	0.22	0.22	206	116	1.05	10087	91.30
22.0	0.27	0.21	208	156	1.41	10243	92.71
22.5	0.32	0.21	209	90	0.81	10333	93.53
23.0	0.36	0.21	211	126	1.14	10459	94.67
23.5	0.41	0.21	213	65	0.59	10524	95.26
24.0	0.45	0.21	214	105	0.95	10629	96.21
24.5	0.50	0.21	216	40	0.36	10669	96.57
25.0	0.54	0.21	218	59	0.53	10728	97.10
25.5	0.59	0.21	219	34	0.31	10762	97.41
26.0	0.63	0.21	221	37	0.33	10799	97.75
26.5	0.68	0.21	222	17	0.15	10816	97.90

Raw Score	Theta	SE	SS	Freq	%	Cum. Freq	Cum. %
27.0	0.73	0.21	224	41	0.37	10857	98.27
27.5	0.77	0.21	225	16	0.14	10873	98.42
28.0	0.82	0.22	227	25	0.23	10898	98.64
28.5	0.87	0.22	228	16	0.14	10914	98.79
29.0	0.91	0.22	230	21	0.19	10935	98.98
29.5	0.96	0.22	231	4	0.04	10939	99.01
30.0	1.01	0.22	233	9	0.08	10948	99.09
30.5	1.06	0.22	234	8	0.07	10956	99.17
31.0	1.11	0.22	236	12	0.11	10968	99.28
31.5	1.16	0.22	237	5	0.05	10973	99.32
32.0	1.21	0.23	238	2	0.02	10975	99.34
32.5	1.26	0.23	240	4	0.04	10979	99.38
33.0	1.31	0.23	241	7	0.06	10986	99.44
33.5	1.36	0.23	243	6	0.05	10992	99.49
34.0	1.42	0.24	244	3	0.03	10995	99.52
34.5	1.48	0.24	245	4	0.04	10999	99.56
35.0	1.53	0.24	246	2	0.02	11001	99.57
35.5	1.59	0.25	248	6	0.05	11007	99.63
36.0	1.65	0.25	249	1	0.01	11008	99.64
36.5	1.72	0.25	250	1	0.01	11009	99.65
37.0	1.78	0.26	251	4	0.04	11013	99.68
37.5	1.85	0.26	253	2	0.02	11015	99.70
38.0	1.92	0.27	254	2	0.02	11017	99.72
38.5	1.99	0.27	255	3	0.03	11020	99.75
39.0	2.07	0.28	256	3	0.03	11023	99.77
39.5	2.15	0.29	257	4	0.04	11027	99.81
40.0	2.24	0.30	258	3	0.03	11030	99.84
40.5	2.33	0.30	260	2	0.02	11032	99.86
41.0	2.42	0.31	261	3	0.03	11035	99.88
41.5	2.52	0.32	262	1	0.01	11036	99.89
42.0	2.63	0.34	263	0	0.00	11036	99.89
42.5	2.75	0.35	264	0	0.00	11036	99.89
43.0	2.88	0.37	265	2	0.02	11038	99.91
43.5	3.02	0.38	266	3	0.03	11041	99.94
44.0	3.17	0.40	267	0	0.00	11041	99.94
44.5	3.35	0.43	268	1	0.01	11042	99.95
45.0	3.54	0.46	269	2	0.02	11044	99.96
45.5	3.77	0.49	270	2	0.02	11046	99.98
46.0	4.04	0.55	271	1	0.01	11047	99.99
46.5	4.37	0.62	272	0	0.00	11047	99.99
47.0	4.83	0.74	273	0	0.00	11047	99.99
47.5	5.58	1.03	274	1	0.01	11048	100.00
48.0	6.82	1.84	300	0	0.00	11048	100.00

Appendix C: Score Calculation Figures and Scoring Rubrics

Table C.1: Score Calculation Figures

Confirmed by NJ DOE (3/8/02)

Used for Means (x)

(Used when 3rd Reader is equal to or adjacent Reader 1 or Reader 2)

Absolute Difference ($ 1^{st} - 2^{nd} $)	Additional Conditions*	Additional Conditions*	Score Calculation*
0	--	--	$(1^{st} + 2^{nd})/2$
1	--	--	$(1^{st} + 2^{nd})/2$
2	$1^{st} < 3^{rd} < 2^{nd}$ or $2^{nd} < 3^{rd} < 1^{st}$	--	$(1^{st} + 2^{nd})/2$
2	$3^{rd} <$ $((1^{st} + 2^{nd})/2)$	$1^{st} < 2^{nd}$	$(1^{st} + 3^{rd})/2$
		$2^{nd} < 1^{st}$	$(2^{nd} + 3^{rd})/2$
	$3^{rd} >$ $((1^{st} + 2^{nd})/2)$	$1^{st} < 2^{nd}$	$(2^{nd} + 3^{rd})/2$
		$2^{nd} < 1^{st}$	$(1^{st} + 3^{rd})/2$
3	$3^{rd} = 1^{st}$ or $(3^{rd} \pm 1) = 1^{st}$	--	$(1^{st} + 3^{rd})/2$
	$3^{rd} = 2^{nd}$ or $(3^{rd} \pm 1) = 2^{nd}$	--	$(2^{nd} + 3^{rd})/2$
4 and 5	$3^{rd} = 1^{st}$ or $(3^{rd} \pm 1) = 1^{st}$	--	$(1^{st} + 3^{rd})/2$
	$3^{rd} = 2^{nd}$ or $(3^{rd} \pm 1) = 2^{nd}$	--	$(2^{nd} + 3^{rd})/2$

Additional Score Calculations

Used for Means (x)

(Used when Reader 3 is NOT equal to or adjacent to Either Reader 1 or Reader 2)

Condition	Score Calculation
$1^{st} < 3^{rd} < 2^{nd}$ or $2^{nd} < 3^{rd} < 1^{st}$	Use 3 rd reading
$1^{st} < 2^{nd} < 3^{rd}$ or $3^{rd} < 2^{nd} < 1^{st}$	$(2^{nd} + 3^{rd})/2$
$2^{nd} < 1^{st} < 3^{rd}$ or $3^{rd} < 1^{st} < 2^{nd}$	$(1^{st} + 3^{rd})/2$

Score Calculation Figure

Confirmed by NJ DOE (3/8/02)

Used for Sum (Σ)

(Used when 3rd Reader is equal to or adjacent Reader 1 or Reader 2)

Absolute Difference ($ 1^{st} - 2^{nd} $)	Additional Conditions*	Additional Conditions*	Score Calculation*
0	--	--	$(1^{st} + 2^{nd})$
1	--	--	$(1^{st} + 2^{nd})$
2	$1^{st} < 3^{rd} < 2^{nd}$ or $2^{nd} < 3^{rd} < 1^{st}$	--	$(1^{st} + 2^{nd})$
2	$3^{rd} <$ $((1^{st} + 2^{nd})/2)$	$1^{st} < 2^{nd}$	$(1^{st} + 3^{rd})$
		$2^{nd} < 1^{st}$	$(2^{nd} + 3^{rd})$
	$3^{rd} >$ $((1^{st} + 2^{nd})/2)$	$1^{st} < 2^{nd}$	$(2^{nd} + 3^{rd})$
		$2^{nd} < 1^{st}$	$(1^{st} + 3^{rd})$
3	$3^{rd} = 1^{st}$ or $(3^{rd} \pm 1) = 1^{st}$	--	$(1^{st} + 3^{rd})$
	$3^{rd} = 2^{nd}$ or $(3^{rd} \pm 1) = 2^{nd}$	--	$(2^{nd} + 3^{rd})$
4 and 5	$3^{rd} = 1^{st}$ or $(3^{rd} \pm 1) = 1^{st}$	--	$(1^{st} + 3^{rd})$
	$3^{rd} = 2^{nd}$ or $(3^{rd} \pm 1) = 2^{nd}$	--	$(2^{nd} + 3^{rd})$

Additional Score Calculations

Used for Sum (Σ)

(Used when Reader 3 is NOT equal to or adjacent to Either Reader 1 or Reader 2)

Condition	Score Calculation
$1^{st} < 3^{rd} < 2^{nd}$ or $2^{nd} < 3^{rd} < 1^{st}$	Use 3 rd reading*2
$1^{st} < 2^{nd} < 3^{rd}$ or $3^{rd} < 2^{nd} < 1^{st}$	$(2^{nd} + 3^{rd})$
$2^{nd} < 1^{st} < 3^{rd}$ or $3^{rd} < 1^{st} < 2^{nd}$	$(1^{st} + 3^{rd})$

Summary of Open-Ended Scoring

Confirmed by NJ DOE (3/8/02)
When to Use the Mean vs. Sum Scoring Rules

Subject	Valid scores	Grade 11
Reading OE	0-4 *	Mean
Writing – Picture	1-6 **	Mean
Writing – Persuasive	1-6 **	Sum
Revise / Edit	0-4 *	Sum
Math OE	0-3 *	Mean
Sci OE	0-3	Mean

* = RF = 6 for Fragment, refusing or unable to write on the topic
 NR = 7 for no response
 OT = 8 for off topic
 NE = 9 for not English

** = NR (No Response=0)
 WF (Wrong Format=7)
 OT (Off Topic=8)
 NE (Not English=9)

Scoring Rubrics

Table C.2: Open-Ended Mathematics Scoring Rubric

3-Point Response

The response shows complete understanding of the problem's essential mathematical concepts. The student executes procedures completely and gives relevant responses to all parts of the task. The response contains few minor errors, if any. The response contains a clear, effective explanation detailing how the problem was solved so that the reader does not need to infer how and why decisions were made.

2-Point Response

The response shows nearly complete understanding of the problem's essential mathematical concepts. The student executes nearly all procedures and gives relevant responses to most parts of the task. The response may have minor errors. The explanation detailing how the problem was solved may not be clear, causing the reader to make some inferences.

1-Point Response

The response shows limited understanding of the problem's essential mathematical concepts. The response and procedures may be incomplete and/or may contain major errors. An incomplete explanation of how the problem was solved may contribute to questions as to how and why decisions were made.

0-Point Response

The response shows insufficient understanding of the problem's essential mathematical concepts. The procedures, if any, contain major errors. There may be no explanation of the solution or the reader may not be able to understand the explanation. The reader may not be able to understand how and why decisions were made.

The above generic rubric is used as a guide to develop specific scoring guides or rubrics for each of the Open-Ended (OE) items which appear on the New Jersey fourth-grade (NJ ASK 4), eighth-grade (GEPA) and eleventh-grade (HSPA) proficiency assessments in mathematics. The generic rubric helps ensure that students are scored in the same way for the same demonstration of knowledge and skills regardless of the test question. More information on Open-Ended items and related scoring is also provided in the *Mathematics Instructional Guide*.

Table C.3: New Jersey Registered Holistic Scoring Rubric

In scoring, consider the grid of written language	Inadequate Command	Limited Command	Partial Command	Adequate Command	Strong Command	Superior Command
Score	1	2	3	4	5	6
Content and Organization	<ul style="list-style-type: none"> May lack opening and/or closing 	<ul style="list-style-type: none"> May lack opening and/or closing 	<ul style="list-style-type: none"> May lack opening and/or closing 	<ul style="list-style-type: none"> Generally has opening and/or closing 	<ul style="list-style-type: none"> Opening and closing 	<ul style="list-style-type: none"> Opening and closing
	<ul style="list-style-type: none"> Minimal response to topic; uncertain focus 	<ul style="list-style-type: none"> Attempts to focus May drift or shift focus 	<ul style="list-style-type: none"> Usually has single focus 	<ul style="list-style-type: none"> Single focus 	<ul style="list-style-type: none"> Single focus Sense of unity and coherence Key ideas developed 	<ul style="list-style-type: none"> Single, distinct focus Unified and coherent Well-developed
	<ul style="list-style-type: none"> No planning evident; disorganized 	<ul style="list-style-type: none"> Attempts organization Few, if any, transitions between ideas 	<ul style="list-style-type: none"> Some lapses or flaws in organization May lack some transitions between ideas 	<ul style="list-style-type: none"> Ideas loosely connected Transitions evident 	<ul style="list-style-type: none"> Logical progression of ideas Moderately fluent Attempts compositional risks 	<ul style="list-style-type: none"> Logical progression of ideas Fluent, cohesive Compositional risks successful
	<ul style="list-style-type: none"> Details random, inappropriate, or barely apparent 	<ul style="list-style-type: none"> Details lack elaboration, i.e., highlight paper 	<ul style="list-style-type: none"> Repetitious details Several unelaborated details 	<ul style="list-style-type: none"> Uneven development of details 	<ul style="list-style-type: none"> Details appropriate and varied 	<ul style="list-style-type: none"> Details effective, vivid, explicit, and/or pertinent
Usage	<ul style="list-style-type: none"> No apparent control Severe/numerous errors 	<ul style="list-style-type: none"> Numerous errors 	<ul style="list-style-type: none"> Errors/patterns of errors may be evident 	<ul style="list-style-type: none"> Some errors that do not interfere with meaning 	<ul style="list-style-type: none"> Few errors 	<ul style="list-style-type: none"> Very few, if any, errors
Sentence Construction	<ul style="list-style-type: none"> Assortment of incomplete and/or incorrect sentences 	<ul style="list-style-type: none"> Excessive monotony/same structure Numerous errors 	<ul style="list-style-type: none"> Little variety in syntax Some errors 	<ul style="list-style-type: none"> Some variety Generally correct 	<ul style="list-style-type: none"> Variety in syntax appropriate and effective Few errors 	<ul style="list-style-type: none"> Precision and/or sophistication Very few, if any, errors
Mechanics	<ul style="list-style-type: none"> Errors so severe they detract from meaning 	<ul style="list-style-type: none"> Numerous serious errors 	<ul style="list-style-type: none"> Patterns of errors evident 	<ul style="list-style-type: none"> No consistent pattern of errors Some errors that do not interfere with meaning 	<ul style="list-style-type: none"> Few errors 	<ul style="list-style-type: none"> Very few, if any, errors

93

NON-SCORABLE RESPONSES	NR = No Response	Student wrote too little to allow a reliable judgment of his/her writing.
	OT = Off Topic/ Off Task	Student did not write on the assigned topic/task, or the student attempted to copy the prompt.
	NE = Not English	Student wrote in a language other than English.
	WF = Wrong Format	Student refused to write on the topic, or the writing task folder was blank.

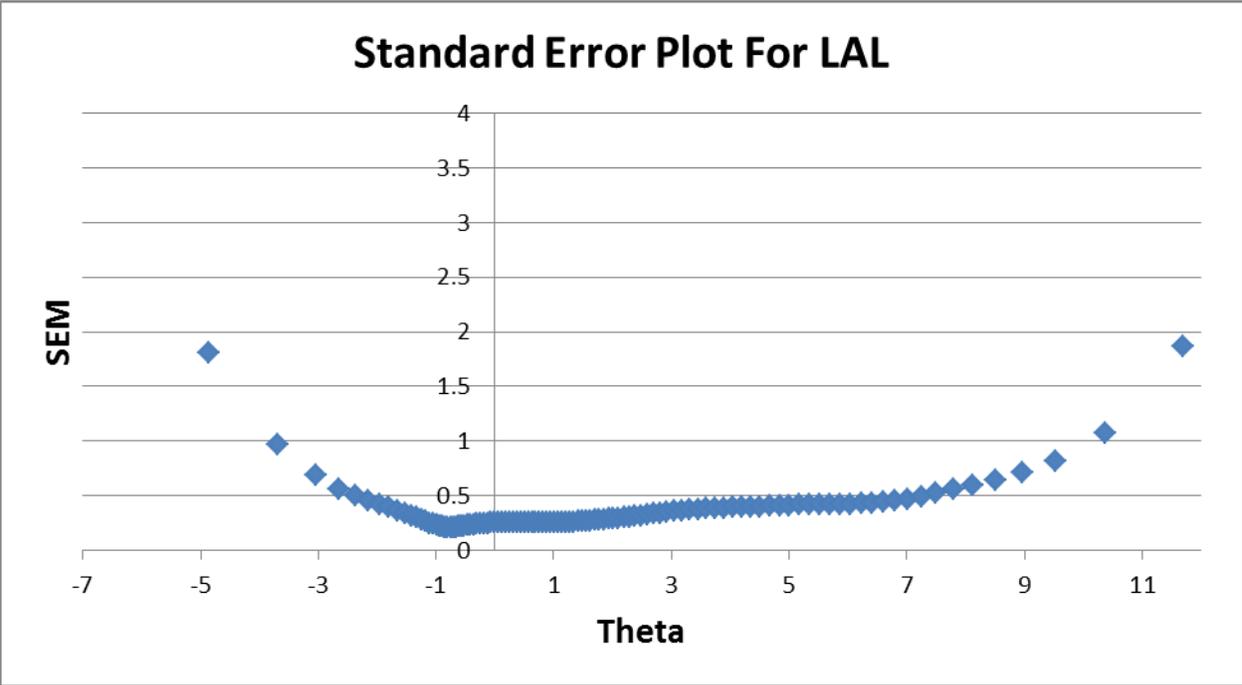
Content/Organization	Usage	Sentence Construction	Mechanics
<ul style="list-style-type: none"> Communicates intended message to intended audience Relates to topic Opening and closing Focused Logical progression of ideas Transitions Appropriate details and information 	<ul style="list-style-type: none"> Tense formation Subject-verb agreement Pronouns usage/agreement Word choice/meaning Proper Modifiers 	<ul style="list-style-type: none"> Variety of type, structure, and length Correct construction 	<ul style="list-style-type: none"> Spelling Capitalization Punctuation

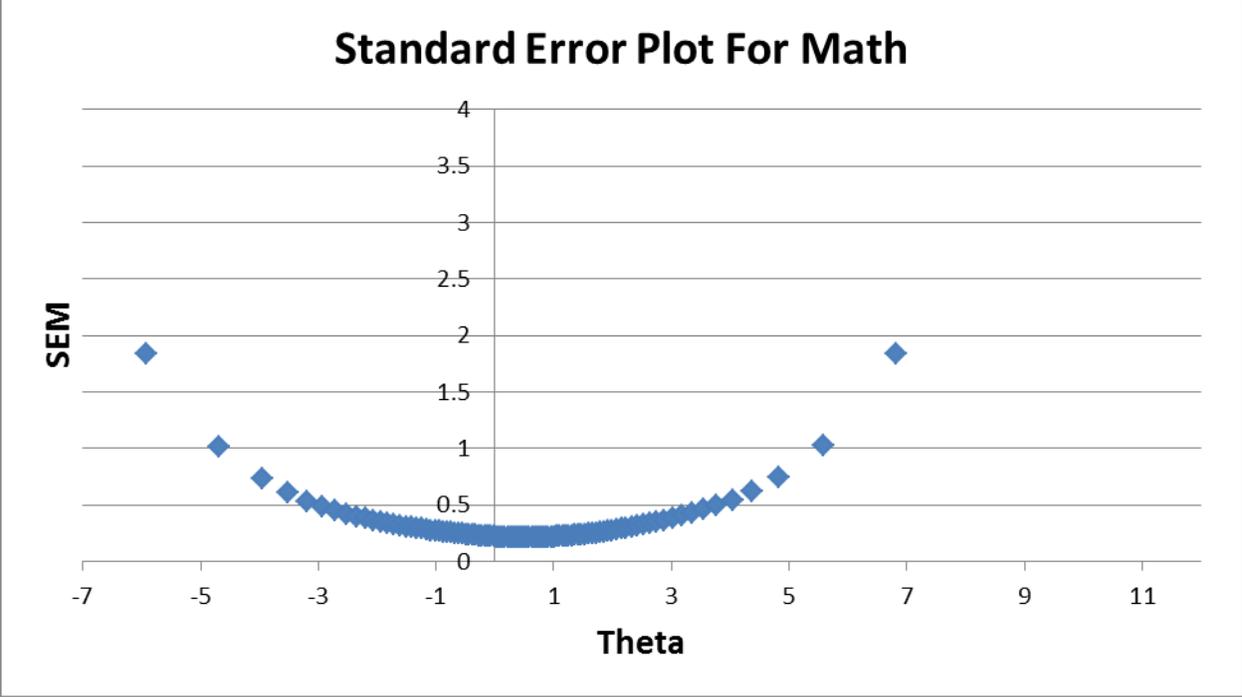
Note: All unscorable responses (NSRs), with the exception of NR, must be coded by the Scoring Director.

Table C.4: Open-Ended Scoring Rubric (Reading)

Points	Criteria
4	A 4-point response clearly demonstrates understanding of the task, completes all requirements, and provides an insightful explanation/opinion that links to or extends aspects of the text.
3	A 3-point response demonstrates an understanding of the task, completes all requirements, and provides some explanation/opinion using situations or ideas from the text as support.
2	A 2-point response may address all of the requirements, but demonstrates a partial understanding of the task, and uses text incorrectly or with limited success resulting in an inconsistent or flawed explanation.
1	A 1-point response demonstrates minimal understanding of the task, does not complete the requirements, and provides only a vague reference to or no use of the text.
0	A 0-point response is irrelevant or off-topic.

Appendix D: Standard Error Plots





Appendix E: Standard Setting Executive Summary

**New Jersey
High School Proficiency Assessment (HSPA)
Standard Setting
June 4-7, 2002**

Executive Summary

Measurement Incorporated

**New Jersey Department of Education
Office of Evaluation and Assessment
Division of Educational Programs and Assessment**

Executive Summary

The New Jersey High School Proficiency Assessment (HSPA) is a statewide, high-stakes assessment and a logical outcome of the New Jersey State Board of Education's adoption of Core Curriculum Content Standards in 1997. The test consists of two sections: a Language Arts Literacy section, a Science Section, and a Mathematics section. The structure and format of the test are summarized in Figure 1. The main goals of the standard-setting process for the HSPA were to (1) describe and delineate the thresholds of performance, for each content area, that are indicative of *Partially Proficient*, *Proficient*, and *Advanced Proficient* performance, and (2) establish recommended cut scores for each content area that differentiate *Partially Proficient* from *Proficient* and *Proficient* from *Advanced Proficient* performance (i.e., two cut scores to yield three performance levels). The recommendations resulting from the standard setting process are designed to help inform the New Jersey State Board of Education as it completes its task of establishing performance standards for the statewide assessments.

From June 4-7, 2002, Measurement Incorporated staff met with representatives of the New Jersey Department of Education (DOE) and 52 educator-panelists from around the state of New Jersey to set performance standards on the HSPA. The panelists, identified by district superintendents, were chosen specifically to represent the demographics of educators throughout the state. Table 1 provides a profile of the 52 panelists, by content area. Following a joint introduction to the process, Language Arts Literacy and Mathematics committees met simultaneously throughout the week. Two senior staff members of Measurement Incorporated, along with two consultants, each with several years of experience in standard setting, led the sessions. Standard setting panelists in each content area were given a thorough orientation and practice test to prepare them for the process and a considerable amount of additional information as they proceeded through three rounds of setting standards, discussing decisions, and settling on final recommendations.

For the Language Arts Literacy component of the New Jersey HSPA standard setting, a holistic sorting method was used. With this method, the panelists' task was to classify student work into one of the three performance categories; the method is holistic in that the panelists considered an individual student's entire body of work – written “open-ended” and multiple-choice items together – in making a classification decision. Proficiency descriptors – that is, general definitions of what a *Proficient* and an *Advanced Proficient* student can do in a particular content area – were developed on April 29, 2002 by another committee of representative New Jersey educators and were provided to the standard-setting panelists. Panelists reviewed folders of actual student papers sampled to represent the full range of scores and were asked to sort these folders into the three performance level categories according to the proficiency descriptors. Ultimately, after having reviewed the implications of their decisions on the distribution of actual student performance in the form of impact data, standard setters identified two cut points to distinguish *Partially Proficient* from *Proficient* performance and *Proficient* from *Advanced Proficient* performance (see Figure 2).

Figure 1. HSPA test structure, March 2002 Base Test

Language Arts Literacy

- 54 total points
- 1 persuasive passage with 10 multiple-choice items (1 point each) and 2 constructed-response items (4 points each)
- 1 narrative passage with 10 multiple choice items (1 point each) and 2 constructed-response items (4 points each)
- 2 writings tasks – 1 picture prompt (6 points) and 1 persuasive writing prompt (12 points)
- 2 content clusters: Working with Text and Analyzing Text
- Day 1: picture prompt (30 minutes), narrative passage and items (50 minutes)
- Day 2: persuasive prompt (60 minutes), persuasive passage and items (45 minutes)

Mathematics

- 48 total points
- 3 30-minute sessions, each with 10 independent multiple-choice items (1 point each) and 2 independent constructed-response items (3 points each)
- 4 content clusters: Number Sense, Concepts, and Applications (12 points); Spatial Sense and Geometry (12 points); Data Analysis, Probability, Statistics, and Discrete Mathematics (12 points); and Patterns, Functions, and Algebra (12 points)

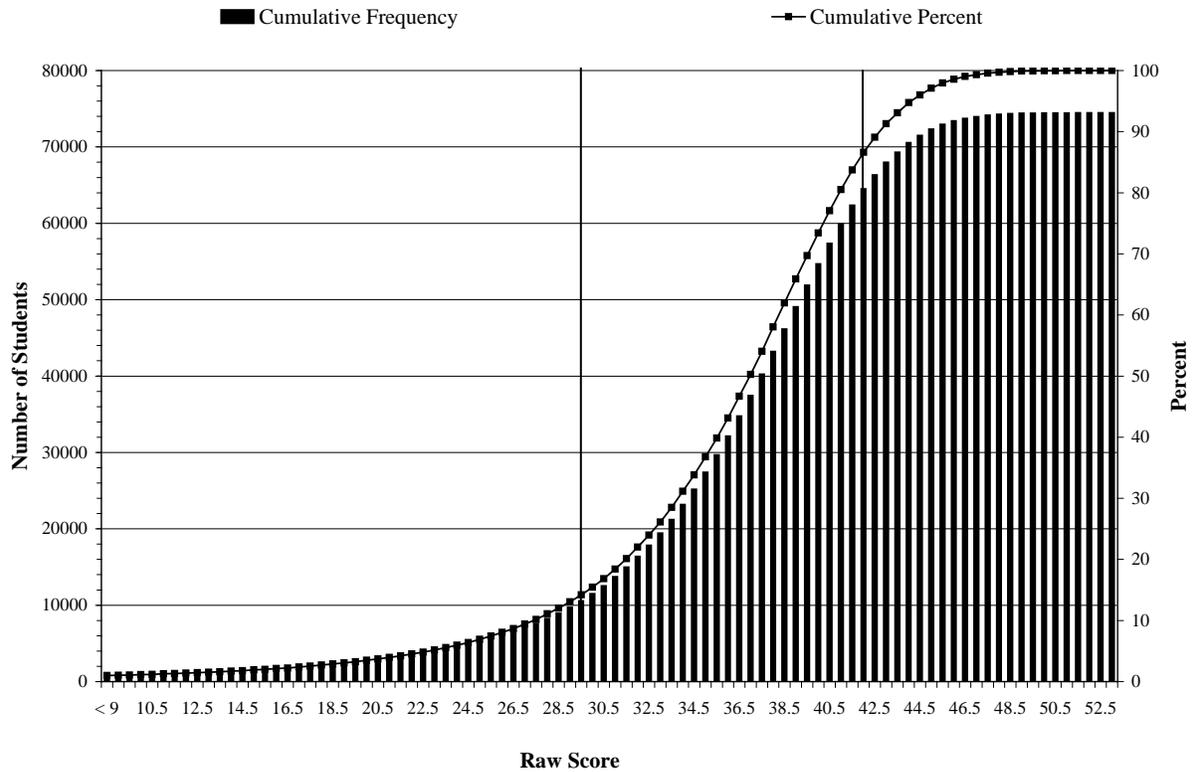
For the Mathematics component of the HSPA, a combination of standard-setting methods was employed. While the holistic sorting method was used to sort students' written responses into the three performance level categories, a Modified Angoff Rating procedure was used to assess the multiple-choice items. Participants were asked to consider the student whose performance can be considered to have just crossed over into the *Proficient* level. They were then asked to estimate, for each multiple-choice item, the probability that this student would answer the item correctly. Likewise, panelists were asked to consider the student whose performance would be considered to have just crossed over into the *Advanced Proficient* level and subsequently estimate the probability that this student would answer the item correctly. Again, panelists were given the opportunity to review impact data before making their final cut score recommendations (see Figure 3).

Based on educator-panelist work, final cut scores for Language Arts Literacy were determined to be 29.5 for *Proficient* and 42.0 for *Advanced Proficient*. As detailed in Figure 2, these performance standards result in the vast majority of students (86.94%) passing the Language Arts Literacy section of the March 2002 HSPA. Final cut scores for Mathematics were determined to be 22.5 for *Proficient* and 39.0 for *Advanced Proficient*. Again, these performance standards result in the majority of students (73.14%) passing the Mathematics section of the March 2002 HSPA (see Figure 3).

Table 1. Panelist Demographic Profile

	Language Arts Literacy	Mathematics
Position	15 teachers	13 teachers
	9 supervisors	13 supervisors/administrators
	2 curriculum coordinators	
Gender	23 females	20 females
	3 males	6 males
Race/ethnicity	22 White	22 White
	4 African-American	2 African-American
		1 Hispanic
		1 Asian
Level of education	5 Doctorate	1 Doctorate
	19 Master's Degree	17 Master's Degree
	2 Bachelor's Degree	8 Bachelor's Degree
DFG representation	3 A	4 A
	3 B	2 B
	3 CD	5 CD
	3 DE	5 DE
	3 FG	2 FG
	2 GH	2 GH
	8 I	5 I
	1 V	1 J
Geographic representation	14 North	11 North
	6 Central	10 Central
	6 South	5 South
Average teaching experience (years)	20.9	25.1
Average teaching experience, 11 th grade (years)	11.6	18.0

Figure 2. Language Arts Literacy Impact Data and Cut Score Recommendations



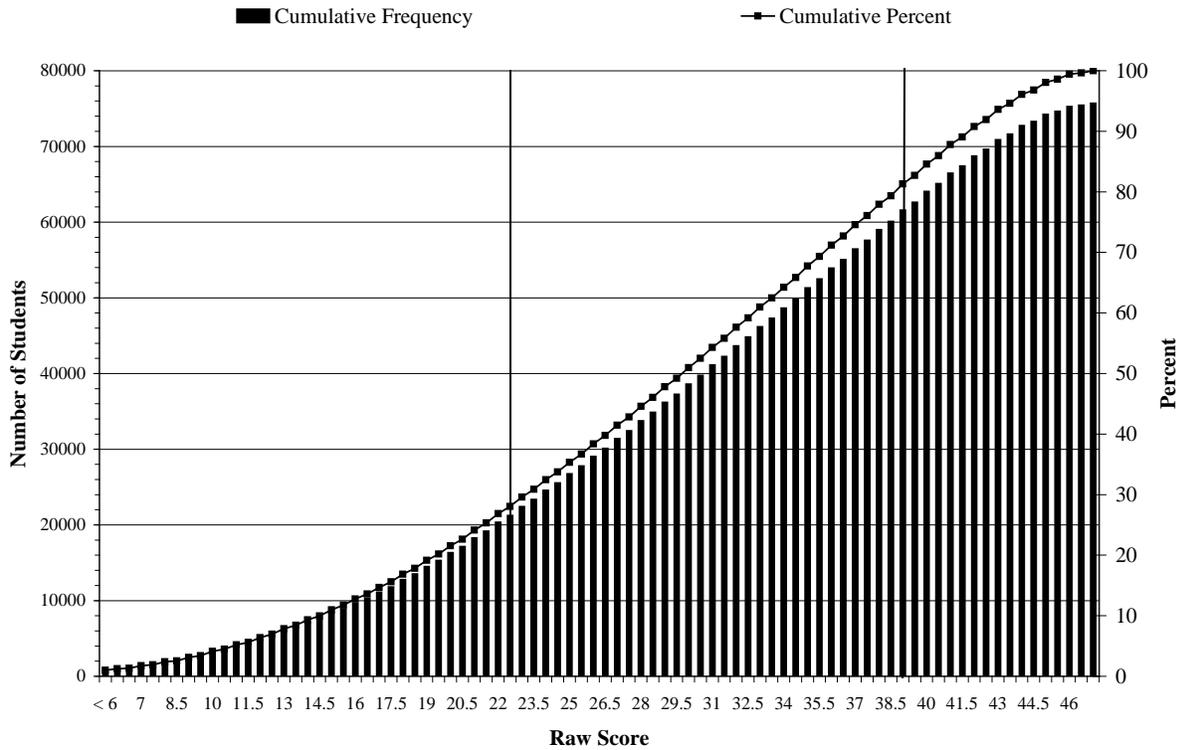
Cut Scores

(Associated standard deviations for the panelist group are shown in parentheses)

	Proficient	Advanced Proficient
Raw score.	29.5 (.39)	42 (.43)
Scaled score	200	250

	Partially Proficient	Proficient	Advanced Proficient
Number of students	9,721	52,629	12,110
Percentage of students taking HSPA	13.06%	70.68%	16.26%

Figure 3. Mathematics Impact Data and Cut Score Recommendations



Cut Scores

(Associated standard deviations for the panelist group are shown in parentheses)

	Proficient	Advanced Proficient
Raw score	22.5 (2.71)	39 (1.70)
Scaled score	200	250

	Partially Proficient	Proficient	Advanced Proficient
Number of students	20,334	39,738	15,628
Percentage of students taking HSPA	26.86%	52.49%	20.64%