Grades 5, 6, and 7
New Jersey Assessment of Skills and Knowledge

TECHNICAL REPORT

NOVEMBER 2007

PTM #1506.48
Copyright © 2007 by New Jersey Department of Education
All rights reserved

NJ ASK 2007 GRADES 5–7 TECHNICAL REPORT
TABLE OF CONTENTS

PART 1: INTRODUCTION.....1

1.1 Description of New Jersey's Overall Assessment System.....1
1.2 Brief Summary of the NJ ASK for Grades 5–7.....1
1.3 Spring Testing and Accountability.....2
1.4 NJ ASK Organizational Support.....2
1.5 Purpose of the Technical Report.....3

PART 2: TEST DEVELOPMENT PROCESS.....4

2.1 Introduction.....4
2.2 Test Specifications.....4
Language Arts Literacy.....4
Mathematics.....5
2.3 Test Blueprints.....6
2.4 Development of Test Items.....11
2.5 Item Review Process.....12
2.6 Test Forms Assembly.....12
2.7 Quality Control for Test Construction.....13
2.8 Summary.....14

PART 3: STANDARD SETTING.....15

3.1 Introduction.....15
3.2 Development of Performance Level Descriptors.....15
3.3 Standard Setting Process.....16
Phase 1.....17
Phase 2.....17
3.4 Summary of Results.....18
3.5 State Board of Education Review and Adoption.....21

PART 4: ITEM ANALYSIS.....22

4.1 Introduction.....22
4.2 Grades 5, 6, and 7 Forms.....22
4.3 Speededness.....29
4.4 Item Bias Statistics.....30
4.5 Summary.....32

PART 5: TEST ADMINISTRATION.....33

5.1 Introduction.....33
5.2 Determining Students for Whom a School Is Accountable.....33
5.3 Accommodations and Modifications.....34
5.4 Testing Exemptions.....34
5.5 Administration of Tests.....35
5.6 Test Security Procedures.....36
5.7 Conclusion.....36

PART 6: SCORING.....37

6.1 Introduction.....37
6.2 Multiple-Choice Items.....37
6.3 Open-Ended Items and Writing Tasks.....37
6.4 Scoring Personnel.....38
6.5 Range Finding and Development of Scoring Guides.....38
6.6 Project Leads.....39
6.7 Training Team Leaders.....39
6.8 Evaluators.....40
6.9 Scoring Procedures.....40
Training Scorers and Qualifying.....40
Monitoring for Quality Assurance.....41
6.10 Conclusion.....42

PART 7: SCALING AND EQUATING.....43

7.1. Introduction.....43
7.2 Scaling and Equating Data.....44
Sample Size and Distributions.....44
Descriptive Statistics for the Equating Samples.....48
7.3. Methods of Scaling and Equating.....51
(1) Re-calibrate the 2006 NJ ASK Test Scores to Create a "Base" Scale.....51
(2) Calibrate 2007 NJ ASK Assessments without Constraint.....52
(3) Examine the Stability of Common Items.....52
(4) Equate the 2007 Assessments to the 2006 Re-calibrated "Base" Scale.....57
(5) Re-center the 2007 Equated Scale to the 2006 Original, or "Reported" Scale.....58
7.4. Summary of Cut Scores for the 2007 NJ ASK.....59
7.5. Methods of Scaling and Equating for Braille and Large Print.....61
Re-calibrate the 2006 NJ ASK Braille and Large Print Tests.....61
Create the 2007 Braille and Larger Print Item Parameter Estimates.....62
Establish Raw Score to Scale Score Tables.....62
7.6. "Special Equating" for Students with Lost Responses.....62

PART 8: REPORTING.....63

8.1 Introduction.....63
8.2 Cycle I Reports.....63
Individual Student Report.....63
Student Sticker.....64
Student Roster.....64
All Sections Roster.....65
Performance Summaries (District and School).....65
8.3 Cycle II Reports.....66
Cluster Means for Students with Valid Scores.....67
Performance by Demographic Groups (Statewide, DFG, District, and School).....68
8.4 State Summary Reporting.....68
8.5 Interpreting Reports.....69

PART 9: ACCOUNTABILITY.....70

9.1 Introduction.....70
9.2 Accountability Model – Overview.....70
9.3 Accountability Model – Goals.....70
9.4 Accountability Classification Results.....71

PART 10: QUALITY CONTROL PROCEDURES.....72

10.1 Quality Control in Data Preparation.....72
10.2 Quality Control in Scanning 72
10.3 Quality Control in Editing and Data Input.....73
10.4 Quality Control in Scoring.....74
10.5 Quality Control in Reporting.....74
10.6 Additional Quality Control Procedures.....75

PART 11: SUMMARY STATISTICS.....76

11.1 Introduction.....76
11.2 Descriptive Statistics for Total Raw Score.....76
11.3 Descriptive Statistics for Total Raw Score by Cluster.....77
11.4 Scale Score Distributions by Content Area and Grade.....78
11.5 Scale Score Distributions by Demographic Group.....78
11.6 Scale Score Distributions by District Factor Group (DFG).....85

PART 12: RELIABILITY.....88

12.1 Introduction.....88
12.2 Reliability and Measurement Error.....88
12.3 Test Metrics and Units of Analysis.....89
12.4 Sources of Measurement Error.....89
12.5 Evidence of Raw Score Internal Consistency.....90
12.6 Evidence Supporting Rater Reliability.....92
12.7 Conditional Estimate of Error at Each Cut-Score.....94
12.8 Reliability of Classifications.....95

PART 13: VALIDITY.....96

13.1 Introduction.....96
13.2 Federal Authority for School Accountability.....96
13.3 Purpose and Intended Uses of Test Performance Scores.....96
13.4 NJ ASK Test Scores.....97
13.5 Content-Related Evidence of Validity.....98
13.6 Appropriateness of Content Definition.....98
13.7 Adequacy of Content Representation.....99
13.8 Validity Evidence Based on the Internal Structure of NJ ASK.....100
13.9 Additional Evidence for Validity of NJ ASK.....102

APPENDICES

APPENDIX 2-1 CHECKLIST FOR FORMS DEVELOPMENT.....103
APPENDIX 3-1 PANELISTS PARTICIPATING IN STANDARD SETTING.....105
APPENDIX 3-2 PERFORMANCE LEVEL DESCRIPTORS.....108
APPENDIX 7-1 RAW SCORE TO SCALE SCORE TABLES.....115
APPENDIX 11-4 SCALE SCORE CUMULATIVE FREQUENCY DISTRIBUTIONS.....137

PART 1: INTRODUCTION

1.1 Description of New Jersey's Overall Assessment System

New Jersey's Office of State Assessments (OSA) coordinates the development and implementation of New Jersey's state-required assessment program designed to measure the extent to which all students at the elementary-, middle-, and secondary-school levels have attained New Jersey's Core Curriculum Content Standards (CCCS). New Jersey's state required assessment system includes the following components:

Elementary School:

- Grade 3 New Jersey Assessment of Skills and Knowledge
- Grade 4 New Jersey Assessment of Skills and Knowledge

Middle School:

- Grade 5 New Jersey Assessment of Skills and Knowledge
- Grade 6 New Jersey Assessment of Skills and Knowledge
- Grade 7 New Jersey Assessment of Skills and Knowledge
- Grade Eight Proficiency Assessment (GEPA)

High School:

- High School Proficiency Assessment (HSPA)

In addition, the statewide assessment program currently includes two tests for special populations:

- Alternate Proficiency Assessment (APA), for students with severe disabilities
- Special Review Assessment (SRA), for students who have not demonstrated proficiency in one or more content areas of the HSPA

1.2 Brief Summary of the NJ ASK for Grades 5–7

This Technical Report focuses on the 2007 New Jersey Assessment of Skills and Knowledge (NJ ASK) administered as an operational test at grades 5, 6, and 7 in the content areas of Language Arts Literacy and Mathematics in spring 2007. The tests provide an indication of student progress toward achieving the knowledge and skills identified in New Jersey's CCCS, and the tests fulfill the requirements under the No Child Left Behind (NCLB) Act.

The NJ ASK is designed to give an early indication of the progress students are making in mastering the knowledge and skills described in the CCCS. The results are to be used by schools and districts to identify strengths and weaknesses in their educational programs. It is anticipated that this process will lead to improved instruction and better alignment with the CCCS. The results may also be used, along with other indicators of student progress, to identify those students who may need instructional support in any of the content areas. This support, which could be in the form of individual or programmatic intervention, would be a means to address any identified knowledge or skill gaps.

The NJ ASK Language Arts Literacy and Mathematics scores at grade 5–7 are reported as scale scores. The scores range from 100–199 (Partially Proficient), 200–249 (Proficient), and 250–300 (Advanced Proficient). The scores of students who are included in the Partially Proficient level are considered to be below the state minimum of proficiency and those students may be most in need of instructional support. The standard-setting procedures used in 2006 for determining proficiency levels are detailed in Part 3 of this Technical Report.

1.3 Spring Testing and Accountability

All testing for NJ ASK is completed in the spring of each year. Recall that the NCLB goal for every school in the state is Proficiency as defined by the New Jersey Board of Education. Testing is conducted in the spring of each year to allow school staff and students the greatest opportunity to achieve the goal of Proficiency.

Data for this Technical Report were collected during the spring administration in April 2007. However, the analyses presented in some Parts (e.g., Part 3 – Standard Setting and Part 7 – Scaling and Equating) of the Report are based on a subset of the total student population. For example, the NJ ASK 2006 grades 5–7 standard setting included student data based on a representative sample of priority districts (available in May) that consisted of about 14% of the student population. Using this representative subset of data from priority districts for standard setting was necessary in order to meet the 2006 reporting timelines established by NCLB. When subsets of data from the total student population are used for analyses, the student N-counts are provided.

Data analyses for the total student population are based on a dataset made available to Riverside Publishing in July 2007. It should be noted that normal reviews of the NJ ASK 2007 grades 5-7 data took place after July and thus there may be very slight differences between the data aggregated in this report and similar data aggregations based on more recent versions of the NJ ASK 2007 dataset.

1.4 NJ ASK Organizational Support

New Jersey's OSA coordinates the development and implementation of the NJ ASK for grades 5–7. In addition to planning, scheduling, and directing all NJ ASK activities, the staff is extensively involved in numerous test reviews, security, and quality-assurance procedures. Riverside is the primary contractor working in partnership with Measurement Incorporated.

The major Riverside activities include program management, development of all test materials (test booklets, answer documents, and ancillary materials), and psychometric support, including standard setting. The major Measurement Incorporated activities include enrollment verification; distribution of all materials; receiving, scanning, editing, and scoring the answer documents; scoring open-ended responses; and creating, generating, and distributing all score reports of test results to students, schools, districts, and the state.

1.5 Purpose of the Technical Report

The purpose of this Technical Report is to provide information about the technical characteristics of the 2007 administration of the NJ ASK for grades 5, 6, and 7. Because the report is technical in nature, with the intended audience being experts in psychometrics and educational research, the report is best understood with a working knowledge of measurement concepts such as reliability and validity, and statistical concepts such as correlation and central tendency. For some chapters, the reader is presumed to have basic familiarity with advanced topics in measurement and statistics such as item response theory (IRT).

This Technical Report provides extensive detail about the development and operation of NJ ASK. The traditional concerns with a program are often labeled reliability and validity. The empirical reliability and validity of the assessments are reported explicitly in this document. While reliability (Part 12) is relatively straightforward, the steps in creating the program and putting it into operation are all aspects of validity (Part 13). The validity of any assessment stems from the steps taken in planning it, the processes of developing the content of the tests, the processes of consulting with stakeholders, the processes of communicating about the test to users, the processes of scoring and reporting, and the processes of data analysis. Each is an inherent part of validity. In short, while there is a specific chapter devoted to validity, this document provides much but not all of the evidence needed to assess the validity of the NJ ASK.

In reading this technical report, it is critical to remember that the testing program does not exist in a vacuum; it is not just a test. It is one part of a complex network intended to help schools focus their energies on dramatic improvement in student learning. NJ ASK is an integrated program of testing, accountability, and curricular and instructional support. It can only be evaluated properly within this full context.

PART 2: TEST DEVELOPMENT PROCESS

2.1 Introduction

The New Jersey Assessment of Skills and Knowledge (NJ ASK) was administered operationally for the second year in 2007 at grades 5, 6, and 7 in the content areas of Language Arts Literacy and Mathematics. The tests provide an indication of student progress toward achieving the knowledge and skills identified in the Core Curriculum Content Standards (CCCS), and the tests fulfill the requirements under the No Child Left Behind (NCLB) Act.

2.2 Test Specifications

Riverside Publishing content experts and the New Jersey Department of Education (NJDOE) developed a directory of test specifications and sample items for each content area. These specifications describe the test, format of the items, and the scores to be generated by the test. This document serves as the foundation for all test item development.

Riverside and the NJDOE rely upon their expertise and the Core Curriculum Content Standards to design a test that is universally accessible to all students in grades 5, 6, and 7 and is composed of test questions that are age- and grade-appropriate. The material in the test specifications is designed for use by curriculum specialists and teachers to improve instruction at the district, school, and classroom levels.

The 2007 NJ ASK administered to students in grades 5, 6, and 7 was designed to measure the same Core Curriculum Content Standards as the NJ ASK administered to students in grades 3 and 4. Brief descriptions of the test content measured in Language Arts Literacy and Mathematics are presented in the following sections.

Language Arts Literacy

The Language Arts Literacy section of each test measures students' achievement in reading and writing. Students read passages selected from published books, newspapers, magazines, and everyday text and respond to related multiple-choice questions, open-ended questions and a writing task.

The Language Arts Literacy assessment currently assesses knowledge and skills in the following clusters (A cluster is a group of related test questions on a single topic.):

- Writing
- Reading
    - Working with Text
    - Analyzing Text

For an in-depth description of the NJ ASK Language Arts Literacy assessment, including specifications, visit the following page on the NJDOE Web site:

http://www.state.nj.us/education/assessment/ms/5-7/specs/

Mathematics

The Mathematics section of each test measures students' ability to solve problems by applying mathematical concepts. The NJ ASK assesses four Core Curriculum Content Standards in Mathematics:

- Number Sense and Numerical Operations
- Geometry and Measurement
- Patterns and Algebra
- Data Analysis, Probability, and Discrete Mathematics

A process cluster, Problem Solving, is also reported on score reports. The process cluster refers to test questions that measure mathematical problem-solving ability. Each test question on the Mathematics assessment measures one content cluster and may contribute to the process cluster. Each content cluster in Mathematics may contain one of the three open-ended items. For an in-depth description of the NJ ASK Mathematics Test Specifications visit the following page on the NJDOE Web site:

http://www.state.nj.us/education/assessment/ms/5-7/specs/

Table 2.2.1
2007 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Total Points Possible by Content Area – Grades 5, 6, and 7

Language Arts Literacy	Grade 5	Grade 6	Grade 7
Total	**41 points**	**48 points**	**48 points**
Writing	5 points	12 points	12 points
Writing/Picture or Speculative	5 points		
Writing/Persuasive Prompt		12 points	12 points
Reading	36 points	36 points	36 points
Working with Text	16 points	15 points	13 points
Analyzing Text	20 points	21 points	23 points
Mathematics	**Grade 5**	**Grade 6**	**Grade 7**
Total	**39 points**	**39 points**	**39 points**
C4.1 - Number Sense & Numerical Operations	10 points	9 points	10 points
C4.2 - Geometry & Measurement	9 points	10 points	9 points
C4.3 - Patterns & Algebra	10 points	10 points	10 points
C4.4 - Data Analysis, Probability, & Discrete Math	10 points	10 points	10 points
Problem Solving	25 points	21 points	36 points

2.3 Test Blueprints

The following tables outline the test construction blueprints. The actual test map for each grade and content area for the 2007 NJ ASK is included.

Table 2.3.1
Test Construction Map for Grade 5 Language Arts Literacy NJ ASK

Text types/Strand	Reading Selections	MC (Number of Items)	OE (Number of Items)	WT (Number of Items)	Total Points
Picture/Speculative Prompt		0	0	1	5
Narrative	1				
AT		4-6	0-2	0	8-12
WT		4-6	0-2	0	8-12
Everyday Text	1				
AT		2-6	0-2	0	8-12
WT		4-8	0-2	0	8-12
Total Items		**20**	**4**	**1**	
Total Points		**20**	**16**	**5**	**41**

Table 2.3.2
Actual Test Map for 2007 Grade 5 Language Arts Literacy NJ ASK

Text types/Strand	Reading Selections	MC (Number of Items)	OE (Number of Items)	WT (Number of Items)	Total Points
Picture/Speculative Prompt		0	0	1	5
Narrative	1				
AT		6	1	0	10
WT		4	1	0	8
Everyday Text	1				
AT		2	2	0	10
WT		8	0	0	8
Total Items		**20**	**4**	**1**	
Total Points		**20**	**16**	**5**	**41**

Table 2.3.3
Test Construction Map for Grade 6 Language Arts Literacy NJ ASK

Text types/Strand	Reading Selections	MC (Number of Items)	OE (Number of Items)	WT (Number of Items)	Total Points
Persuasive Prompt		0	0	1	12
Narrative	1				
AT		4-6	0-2	0	6-12
WT		4-6	0-2	0	6-12
Everyday Text	1				
AT		4-6	0-2	0	6-12
WT		4-6	0-2	0	6-12
Total Items		**20**	**4**	**1**	
Total Points		**20**	**16**	**12**	**48**

Table 2.3.4
Actual Test Map for 2007 Grade 6 Language Arts Literacy NJ ASK

Text types/Strand	Reading Selections	MC (Number of Items)	OE (Number of Items)	WT (Number of Items)	Total Points
Persuasive Prompt		0	0	1	12
Narrative	1				
AT		5	1	0	9
WT		5	1	0	9
Everyday Text	1				
AT		4	2	0	12
WT		6	0	0	6
Total Items		**20**	**4**	**1**	
Total Points		**20**	**16**	**12**	**48**

Table 2.3.5
Test Construction Map for Grade 7 Language Arts Literacy NJ ASK

Text types/Strand	Reading Selections	MC (Number of Items)	OE (Number of Items)	WT (Number of Items)	Total Points
Persuasive Prompt		0	0	1	12
Narrative	1				
AT		4-6	0-2	0	6-12
WT		4-6	0-2	0	6-12
Everyday Text	1				
AT		4-6	0-2	0	6-12
WT		4-6	0-2	0	6-12
Total Items		**20**	**4**	**1**	
Total Points		**20**	**16**	**12**	**48**

Table 2.3.6
Actual Test Map for 2007 Grade 7 Language Arts Literacy NJ ASK

Text types/Strand	Reading Selections	MC (Number of Items)	OE (Number of Items)	WT (Number of Items)	Total Points
Persuasive Prompt		0	0	1	12
Narrative	1				
AT		4	2	0	12
WT		6	0	0	6
Everyday Text	1				
AT		3	2	0	11
WT		7	0	0	7
Total Items		**20**	**4**	**1**	
Total Points		**20**	**16**	**12**	**48**

Table 2.3.7
Test Construction Map for Grade 5 Mathematics NJ ASK

Standard	Specified MC	Actual (1 pt.)	Specified OE	Actual (3 pts.)	Total Items	Total Points
I	6–10		0–1			
II	6–10		0–1			
III	6–10		0–1			
IV	6–10		0–1			
Total Items	**30**		**3**		**33**	
Total Points	**30**		**9**			**39**

Table 2.3.8
Actual Test Map for 2007 Grade 5 Mathematics NJ ASK

Standard	Specified MC	Actual (1 pt.)	Specified OE	Actual (3 pts.)	Total Items	Total Points
I	6–10	7	0–1	1	8	10
II	6–10	9	0–1	0	9	9
III	6–10	7	0–1	1	8	10
IV	6–10	7	0–1	1	8	10
Total Items	**30**	**30**	**3**	**3**	**33**	
Total Points	**30**	**30**	**9**	**9**		**39**

Table 2.3.9
Test Construction Map for Grade 6 Mathematics NJ ASK

Standard	Specified MC	Actual (1 pt.)	Specified OE	Actual (3 pts.)	Total Items	Total Points
I	6–10		0–1			
II	6–10		0–1			
III	6–10		0–1			
IV	6–10		0–1			
Total Items	**30**		**3**		**33**	
Total Points	**30**		**9**			**39**

Table 2.3.10
Actual Test Map for 2007 Grade 6 Mathematics NJ ASK

Standard	Specified MC	Actual (1 pt.)	Specified OE	Actual (3 pts.)	Total Items	Total Points
I	6–10	9	0–1	0	9	9
II	6–10	7	0–1	1	8	10
III	6–10	7	0–1	1	8	10
IV	6–10	7	0–1	1	8	10
Total Items	**30**	**30**	**3**	**3**	**33**	
Total Points	**30**	**30**	**9**	**9**		**39**

Table 2.3.11
Test Construction Map for NJ ASK Grade 7 Mathematics NJ ASK

Standard	Specified MC	Actual (1 pt.)	Specified OE	Actual (3 pts.)	Total Items	Total Points
I	6–10		0–1			
II	6–10		0–1			
III	6–10		0–1			
IV	6–10		0–1			
Total Items	**30**		**3**		**33**	
Total Points	**30**		**9**			**39**

Table 2.3.12
Actual Test Map for 2007 NJ ASK Grade 7 Mathematics NJ ASK

Standard	Specified MC	Actual (1 pt)	Specified OE	Actual (3 pts.)	Total Items	Total Points
I	6-10	7	0-1	1	8	10
II	6-10	9	0-1	0	9	9
III	6-10	7	0-1	1	8	10
IV	6-10	7	0-1	1	8	10
Total Items	**30**	**30**	**3**	**3**	**33**	
Total Points	**30**	**30**	**9**	**9**		**39**

2.4 Development of Test Items

The 2007 Grade 5, 6, and 7 NJ ASK consists of operational or base test items used to determine students' scores. These items originated from three sources:

1. Items field-tested on a former New Jersey Proficiency Assessment of State Standards (NJPASS) test but never used operationally. All Language Arts Literacy tests were constructed from this pool of items. Approximately one-half of the items used to construct the Mathematics tests were selected from this pool of items. All items were aligned to the current New Jersey Core Curriculum Content Standards.
2. Items field-tested with students from the state of Georgia and may have been used operationally in a secure high-stakes testing environment. Approximately one-half of the items used to construct the Mathematics tests were selected from this pool of items. All items were aligned to the current New Jersey Core Curriculum Content Standards.
3. Items field-tested with Ohio students that appeared on the Riverside product, OOPT 2nd Edition. There were two items pulled from this pool. Both items were slightly modified with Riverside psychometric approval to meet current NJ requirements and aligned to the current New Jersey Core Curriculum Content Standards.

All items used on the 2007 NJ ASK went through the following steps of the item development process prior to 2000.

1. Riverside: Created test and item specifications based on requirements of state
2. Riverside: Selected and trained item writers
3. Item Writers: Wrote test items
4. Riverside: Conducted initial item review
5. Riverside: Conducted item review by experienced senior staff
6. Riverside: Conducted content and bias review with committees comprised of educators.

7. Field-tested items with New Jersey students (1998), with Ohio students, or with Georgia students.
8. Riverside: Conducted Statistical Item Review

In December 2005, January 2006, and Fall of 2006 the following additional development processes were undertaken.

9. Riverside: Aligned items to the New Jersey Core Curriculum Content Standards
10. NJDOE: Approved alignment of items, including the balance of standards reflected in the test blueprint; also improved item quality.
11. Removed all items that did not have NJDOE approval for adherence to the New Jersey Core Curriculum Content Standards.

2.5 Item Review Process

Riverside brought all potential operational items to New Jersey. Riverside reviewed the items with the NJDOE for strict adherence to the New Jersey Core Content Curriculum Standards. Any item that did not strictly adhere to an alignment was removed and replaced. Items were also reviewed for quality with respect to construct standards and data. Riverside and the NJDOE attempted to ensure that no changes were made to the items so that the data would not be affected. If the item did not meet the construct standard or was determined to have poor data, it was removed and replaced. No item used operationally was revised in such a way that it would have changed the alignment.

2.6 Test Forms Assembly

There were four steps associated with assembling test forms for NJ ASK:

1. Determine form design
2. Select items that meet content specifications
3. Evaluate statistical specifications and select items to meet these specifications
4. Review and approve test forms

Determine forms design – Each form consists of a set of operational items.

Select items that meet content specifications – Each content area contains subsets of items called clusters. In Language Arts Literacy the clusters include: Writing (Writing about Pictures or Writing about a Speculative prompt [grade 5] and Writing about Persuasive Prompts [grades 6 and 7]) and Reading (Working with Text and Analyzing Text). In Mathematics the clusters include Number Sense and Numerical Operations; Geometry and Measurement; Patterns and Algebra; and Data Analysis, Probability, and Discrete Mathematics. There is also a process cluster called Problem Solving.

Evaluate statistical specifications and select items to meet these specifications – As forms are created it is necessary to determine if the statistical specifications have been met. Spreadsheets (form matrices) are used to provide information on the statistical properties of newly created forms. These matrices contained the following statistics (among others): Average p-values, point-biserial correlations and IRT difficulty values. Riverside's research group and the NJDOE each conducted an internal review of the data to ensure equivalence of test difficulty across years. For example, a set of anchor items across adjacent years remained constant so the 2007 test forms could be equated to the 2006 test forms.

Review and approve test forms – Once the content and statistical specifications have been met for each grade and subject, the forms are approved by the NJDOE. The forms are then released for production and editorial reviews.

Checklists and quality control procedures accompany each stage of forms development. Some of these procedures are listed below.

2.7 Quality Control for Test Construction

Following is a list of quality control procedures used during the assembly of NJ ASK forms:

- Construct forms based on all content requirements noted in the test blueprint.
- Verify correct number of items per standard or reporting category based on test blueprint.
- Review selected items to ensure a wide sampling of the knowledge and skills being measured.
- Ensure that all selected items have been through the appropriate review procedures and are approved for use by the NJDOE.
- Check for a variety of item topics, equal distribution of male/female, ethnicities, etc.
- Verify appropriate portions of items with and without artwork.
- Check for cueing across all items on each form.
- Verify match of unique item identification numbers (UIN) to test matrix.
- Verify equal or nearly equal distribution of answer choices for MC items.
- Ensure that the test meets the statistical specifications, i.e., that as many items as possible have p-values between .35 and .9; as many items as possible have point-biserials above .20; and the average Rasch value is between 495 and 500 (see Appendix 2-1).
- Consider any statistical flags or problems.
- Check statistics to ensure that the collection of items yields an overall difficulty that falls within the specified range.
- Verify that items have not been released to the public.
- Verify correct answer key for each item.
- Perform content review of form (senior staff).
- Perform statistical review of form (psychometrician/statistician).
- Send form to NJDOE for review and approval.

2.8 Summary

The NJ ASK for grades 5–7 provides an indication of student progress toward achieving the knowledge and skills identified in the Core Curriculum Content Standards (CCCS) and the tests fulfill the requirements under NCLB. Just as the CCCS guided the item development and selection process, the consideration of content played an equally important role in forms development. Forms development required a balance of both content coverage and item difficulty. As items were selected for inclusion on particular forms, every effort was made to balance the content coverage to ensure they aligned to the CCCS being assessed while simultaneously considering the overall difficulty of the forms.

PART 3: STANDARD SETTING

3.1 Introduction

Both the 2006 and 2007 NJ ASK assessed two subject areas, Language Arts Literacy (LAL) and mathematics, in grades 5, 6, and 7. Each subject area included both multiple-choice items and open-ended items. After the first administration of the NJ ASK in April 2006, a standard setting workshop was held to determine the cut scores for LAL and mathematics that would distinguish performance among three levels: Partially Proficient, Proficient, and Advanced Proficient. Because the cut scores established in 2006 also apply to the 2007 NJ ASK, a detailed description of the 2006 standard setting is provided in this section.

The NJ ASK 2006 grades 5–7 standard setting had two phases. Phase 1 was a two-day session and was held on May 11 and 12, 2006. It involved 118 educators from across the state of New Jersey (see Appendix 3-1 for the demographic background of panelists) and used a research-based standard setting method to recommend cut scores. The Phase 2 meeting was held on May 15, 2006. It included one person from each of the Phase 1 panels as well as five additional policymakers from the state. A state board meeting was held on May 17, 2006, in Trenton, and the recommended cut scores from Phase 2 were approved.

The full Standard Setting report, available from the New Jersey Department of Education (NJDOE), provides details about the standard setting procedures, demographic information of the panelists, panelists' ratings from one round to the next, and their responses on the evaluation forms. The final cut scores approved by the State Board of Education are also presented. The sections below summarize the most important steps of the standard setting process. For more detail, the full Standard Setting Report should be referenced.

3.2 Development of Performance Level Descriptors

On April 19, 2006, a group of 57 educators participated in a one-day workshop to develop the Performance Level Descriptors (PLDs) for Proficient and Advanced Proficient in LAL and mathematics. These descriptors were used throughout the standard-setting process and can found in Appendix 3-2. The NJ ASK 2006 grades 5–7 PLD meeting included the following steps:

1. Orientation - All panelists gathered as one large group to view a Power Point presentation prepared by the NJDOE. The presentation discussed the development of NJ ASK 2006 grades 5–7 and apprised the panelists as to what a PLD is and how it would be used by panelists in Phase 1 of standard setting to develop cut scores. For the day's activities each panelist was given a copy of New Jersey's Core Curriculum Content Standards (CCCS).
2. Panelists then broke up into six preplanned groups - one per content area and grade level. Each panelist then took the appropriate NJ ASK 2006 grades 5–7 operational test.

3. Panelists were provided time to discuss the test, including its content, challenges, and difficulty.
4. Panelists divided into smaller groups and made lists of cognitive characteristics required by the NJ CCCS that would define Proficient and Advanced Proficient for their assigned grade and content area. This process was facilitated by NJDOE staff, and the documents and the prompts were supplied by the NJDOE.
5. All panelists returned to their grade and content group and the cognitive characteristics developed in Step 4 were merged into a narrative PLD. This process was facilitated by NJDOE staff, and the documents and the prompts were supplied by the NJDOE.
6. PLDs from all three grade's PLDs were read to all of the mathematics and language arts literacy panelists, and time was allowed for edits and suggestions from content appropriate panelists from other grades and NJDOE facilitators. PLDs were checked for vertical scaling during this time.
7. Panelists returned to grade assignments to consider and incorporate edits and suggestions to their PLDs.
8. Final PLDs were then presented to all appropriate content panelists.
9. Panelists completed a questionnaire developed by NJDOE.
10. PLDs were presented to all members of the Office of Evaluation and Assessment for edits and suggestions. No edits or suggestions were made and the PLDs that were developed on April 19, 2006, were used on May 11 and 12 for Standard Setting Phase 1.

3.3 Standard Setting Process

The NJ ASK 2006 grades 5–7 standard setting included student data based on a sample of data from priority districts that consisted of about 14% of the student population. The scores for 42,569 students from grades 5, 6, and 7 were used in setting the LAL standards, and a total of 42,592 students were used in setting the mathematics standards.

Since the LAL and mathematics tests included both multiple-choice and open-ended items, a modified-Angoff method (Hambleton & Plake, 1995)[1] was implemented. The Angoff method is the most thoroughly researched method used in setting standards. According to Loomis & Bourke (2001)[2], "No other method could be identified that appeared to be as easy to use, as technically sound, and as well researched as the Angoff method."

[1] Hambleton, R. K. & Plake, B. S (1995) Using an Extended Angoff Procedure to Set Standards on Complex Performance Assessments. *Applied Measurement in Education*, **8**, 1, 41-55.
[2] Loomis, S. C. & Bourque, M. L. (2001). From tradition to innovation: Standard setting on the National Assessment of Educational Progress. In G. J. Cizek (Ed.), Standard performance standards: Concepts, methods, and perspectives. Mahwah, NJ: Lawrence Erlbaum Associates.

Phase 1

The NJ ASK 2006 grades 5–7 standard-setting procedures contained two phases. Phase 1 included the following 10 steps:

1. Orientation to the NJ ASK 2006 program and the purpose and use of standards.
2. Each judge taking the test.
3. Discussing and operationally defining the *Proficient* and *Advanced Proficient* Performance Level Descriptors.
4. Initial, independent ratings by judges, each estimating the number of students out of 100 who would answer each multiple-choice test question correctly. Judges were provided forms (e.g., Standard-Setting Item Rating Form) for accomplishing this task.
5. Data tabulation for each judge, and across averages across all judges, to derive the mean cut score for the panel.
6. Discussion of initial estimates derived from the Stage 1 ratings and review of item-level information (p-values) based on actual student data from the priority districts sample.
7. Stage 2 ratings occurred after reviewing results and discussing the new information. Each judge had an opportunity to change or reaffirm his/her decision regarding each item.
8. Discussion of the overall judges' Stage 2 cut scores for the two standards of *Proficient* and *Advanced Proficient*.
9. Final estimates were determined after presentation and discussion of the Stage 2 data. Based on these Stage 3 ratings, recommendations were made for *Proficient* and *Advanced Proficient* performance.
10. Each judge was asked to complete an evaluation of the standard-setting process.

Phase 2

The Phase 2 meeting took place on May 15, 2006, three days after the modified-Angoff event. One judge from each of the Phase 1 panels and five policymakers, including some superintendents and other stakeholders, participated in this meeting.

The purpose of the Phase 2 panel was to look at the big picture of the standards across grades and subjects. Their responsibility was to review the Phase 1 results from a policy point of view and make any changes that were necessary from this viewpoint. The Phase 2 panelists reviewed the results and considered consistency across all grades (grades 3 through 8) and assessed content areas in terms of both the percentage of points required to reach each performance level and the percentage of students reaching each performance level. They were provided information on how the initial cut scores were determined, the value of the cut scores, and the projected impact in the student population. Phase 2 panelists were also given graphs which showed the percentage of students in each classification group for Grades 3, 4,

and 8, as well as the percentage of students in each classification group developed by the Grade 5–7 standard setting.

The primary task of Phase 2 panelists was to determine the reasonableness of the standards from a policy perspective and to recommend adopting either the Phase 1 cut scores or another set of cut scores within reasonable limits of the originals (e.g., up to +/- 4-point range).

Phase 2 of NJ ASK 2006 grades 5–7 standard setting included the following:

1. Choose the Panel. The panel was composed of six individuals consisting of a subset of the modified-Angoff panelists supplemented by five state and local policymakers, including district superintendents and curriculum supervisors. The modified-Angoff panelists were recommended from each of the six panels. An attempt was made to select modified-Angoff panelists who the NJDOE and RPC felt could best represent their respective panel in terms of having the ability to communicate the groups' thoughts and discussions. In addition, modified-Angoff panelists provided representation from both genders and various ethnic and DFG groups.
2. Summarize Phase 1 Procedure. It was important to ensure that the Phase 2 panelists had a complete understanding of the Phase 1 process. To this end, facilitators described the modified-Angoff procedures to the Phase 2 panelists so they could understand how the cut scores were determined. Participants from Phase 1 were asked to provide their own insights on important issues discussed in determining the Phase 1 cut scores.
3. Present Impact Data. The Phase 2 panelists were shown tabled results from the Phase 1 event, including impact data for the recommended cut scores. In addition, impact data for cut scores +/- 2 raw score points were also provided. Impact data were provided for all students as well as for important state and NCLB reporting groups, such as gender, ethnic group, educational programs, economical status, and migrant status. Panelists were also asked to consider the importance of consistency across grades and subjects.
4. Discuss Final Cut Scores. After panelists were trained on the methodology and saw the data described above, they discussed the results. Facilitators focused the panelists' discussion on whether the percentage of students reaching Proficient and Advanced Proficient were viable percentages in New Jersey.
5. Make Final Recommendations. After reviewing the Proficient and Advanced Proficient impact data and discussing the percentages both within and across grades levels, the Phase 2 panelists came to a decision on what cut scores to recommend. Cut scores recommended by Phase 2 panelists were submitted to the New Jersey State Board of Education.

3.4 Summary of Results

Overall, panelists' judgments about the cut scores converged from one round to the next, showing little variance by the end of the final round of Phase 1. This was reflected in the

standard error of judgment (SEJ) that is related to the variance in panelists' judgments around the cut score. More specifically, the SEJ is the standard deviation calculated from the total score based on the sum of the judges' p-value ratings across all test items. In general, the SEJs decreased across rounds, indicating converging opinions. See the RESULTS section of the full Standard Setting Report (page 27) for more details.

The NJ ASK 2006 grades 5, 6, and 7 cut scores and the corresponding conditional standard error of measurement (CSEM) approved by the New Jersey State Board of Education on May 17, 2006, are summarized in Table 3.4.1. The CSEM is calculated as:

$$CSEM = (SD_{obs} / SD_{adj'd}) * SE_{theta}$$

Where: SD_{obs} is the observed standard deviation for number correct raw score,
$SD_{adj'd}$ is an estimate of the "true" sample standard deviation, and
SE_{theta} is the standard error for theta.

Table 3.4.1
NJ ASK 2006 Grades 5–7 Cut Scores with Conditional Standard Error of Measurement by Content Area and Grade Level

		Language Arts Literacy		Mathematics	
		Proficient	Advanced Proficient	Proficient	Advanced Proficient
Grade 5	Cut score*	16	30	18	30
	CSEM	1.7	2.7	2.5	2.7
Grade 6	Cut score*	20	32	17	31
	CSEM	1.5	2.0	3.8	4.4
Grade 7	Cut score*	21	34	13	26
	CSEM	1.5	2.2	3.4	3.2

*Cut scores were approved by the New Jersey State Board of Education on May 17, 2006.

Distributions of students in each proficient category—Partially Proficient (PP), Proficient (P), and Advanced Proficient (AP)—are tabulated in Table 3.4.2 and presented graphically in Figure 3.4.1.

Table 3.4.2
Distribution of Students in Each Proficient Category and Test Characteristics by Content Area and Grade Level

	Grade	N-count*	% Partially Proficient	% Proficient	% Advanced Proficient	N-item	Mean Raw Score	Maximum Possible Score
LAL	5	14181	21.2	70.7	8.1	25	20.80	41
	6	14206	36.1	56.5	7.4	25	21.64	48
	7	14182	37.5	56.8	5.7	25	22.42	48
Math	5	14185	21.4	56.0	22.6	33	23.41	39
	6	14208	33.6	53.0	13.4	33	20.53	39
	7	14199	40.6	47.6	11.8	33	15.57	39

*Based on data from priority districts.

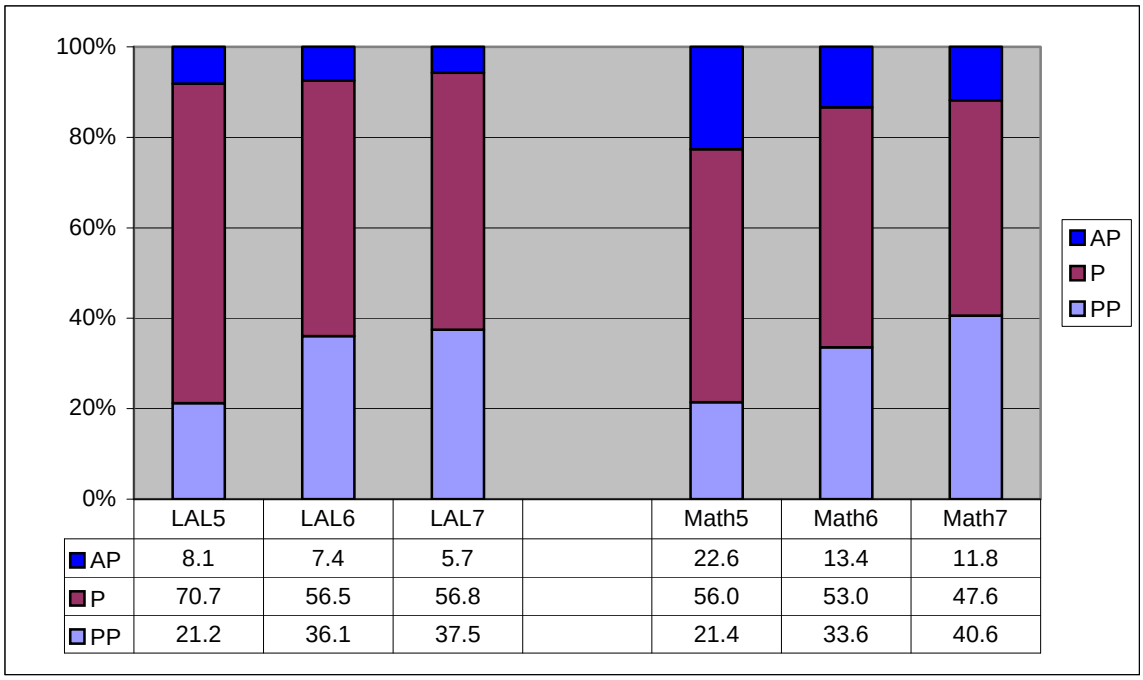

Figure 3.4.1: Distribution of Students in Each Proficient Category by Content Area and Grade Level

3.5 State Board of Education Review and Adoption

A New Jersey State Board of Education meeting was held May 17, 2006, two days after the Phase 2 meeting. An introduction was first presented to the Board members. The introduction included a definition of the concept of "Standard Setting," NJ ASK proficiency levels, the Angoff procedure, and validity considerations. Following that, the methods of rating and calculating the cut scores were demonstrated. Phase 1 and Phase 2 results were presented. In addition, distributions of students by subgroups, such as gender, ethnic group, and educational programs, were presented. Validity evidence of the standard setting procedures as well as the cut scores were also presented. All cut scores recommended by the Phase 2 panelists were approved by the State Board of Education on May 17, 2006. As noted in the Introduction of this Section, the cut scores established in 2006 also apply to the 2007 NJ ASK.

PART 4: ITEM ANALYSIS

4.1 Introduction

Item analyses were conducted for the 2007 NJ ASK for grades 5, 6, and 7 in the content areas of Language Arts Literacy (LAL) and Mathematics. In this section, we present summary information by grade at both the content domain and content cluster level. The information includes mean item scores and discrimination indices. The data summarized in this section are based on the 2007 operational administration and include responses from 101,941, 103,689, and 104,902 students, for grades 5, 6, and 7, respectively.

For multiple-choice (MC) items, the mean score is simply the proportion of students who gave a correct response to the item (usually referred to as item difficulty or the p-value), and the discrimination index is the point biserial correlation between the item score and the total score based on the remaining items. For LAL, the test criterion score was the total score of all reading items (MC and OE) and the writing prompt. For mathematics, the test criterion score was the total score of all MC and OE items.

For open-ended (OE) items, the mean score is the mean of students' scores on a scale of 0 to 3 for mathematics and 0 to 4 for LAL. Writing is scored on a scale of 0 to 5 for grade 5 and 0 to 6 for grades 6 and 7. Note that for grades 6 and 7, the writing scores were doubled in data analyses and score reporting. The discrimination index is the correlation between the item score and the total score based on the remaining items.

4.2 Grade 5, 6 and 7 Forms

Tables 4.2.1, 4.2.2, and 4.2.3 summarize, by item response format, item difficulty and discrimination of the items that comprise each content domain and cluster for grades 5, 6, and 7 respectively. For MC items, both the mean and standard deviation are given. The mean value is the average of the p-values of the items in the cluster. For OE items, the mean value is the average item score for the items in the cluster. Item discrimination is the correlation between students' item score and the total score of the remaining items on the test. Both item difficulty and discrimination are expressed in terms of the raw score metric.

Tables 4.2.4, 4.2.5, and 4.2.6 summarize frequency distributions for MC item difficulty and discrimination indices of items comprising each content domain and cluster for grades 5, 6, and 7, respectively. The median item difficulty and discrimination is also displayed.

Table 4.2.7 summarizes distractor analyses for MC items by test form. The number in each cell indicates the number of items where at least one p-value or discrimination index (point-biserial) for the distractors was higher than the keyed option.

Table 4.2.1
Grade 5 - Item Difficulty and Discrimination Summary Statistics for Multiple-Choice and Open-Ended Items by Content Area and Cluster

Test Section/ Cluster	Multiple-choice				Open-ended*			
	Item Difficulty			Item Discrimination	Item Difficulty			Item Discrimination
	Nitem	Mean	S.D.	Mean	Nitem	Mean	S.D.	Mean
LAL	**20**	**0.77**	**0.07**	**0.42**				
Writing					1	2.76		0.56
Reading	20	0.77	0.07	0.42	4	1.37	0.32	0.51
Interpreting Text	12	0.75	0.08	0.40	1	1.66		0.58
Analyzing Text	8	0.81	0.05	0.45	3	1.28	0.32	0.49
Math	**30**	**0.69**	**0.17**	**0.34**	**3**	**1.12**	**0.27**	**0.47**
Number and Numerical Operation	7	0.77	0.09	0.40	1	1.28		0.55
Geometry and Measurement	9	0.69	0.21	0.28	0			
Patterns and Algebra	7	0.69	0.18	0.33	1	1.27		0.37
Data Analysis, Probability, and Discrete Mathematics	7	0.63	0.17	0.38	1	0.80		0.50
Problem Solving	16	0.74	0.26	0.35	3	1.12	0.27	0.47

*For open-ended (OE) items, the mean score is the mean of students' scores on a scale of 0 to 3 for mathematics and 0 to 4 for LAL. Writing is scored on a scale of 0 to 5 for grade 5 and 0 to 6 for grades 6 and 7. Note that for grades 6 and 7, the writing scores were doubled in data analyses and score reporting.

Table 4.2.2
Grade 6 - Item Difficulty and Discrimination Summary Statistics for Multiple-Choice and Open-Ended Items by Content Area and Cluster

Test Section/ Cluster	Multiple-choice				Open-ended*			
	Item Difficulty			Item Discrimination	Item Difficulty			Item Discrimination
	Nitem	Mean	S.D.	Mean	Nitem	Mean	S.D.	Mean
LAL	**20**	**0.67**	**0.13**	**0.37**				
Writing					1	5.85		0.60
Reading	20	0.67	0.13	0.37	4	1.27	0.16	0.53
Interpreting Text	11	0.60	0.12	0.34	1	1.17		0.51
Analyzing Text	9	0.76	0.09	0.40	3	1.30	0.18	0.54
Math	**30**	**0.60**	**0.19**	**0.36**	**3**	**1.29**	**0.50**	**0.58**
Number and Numerical Operation	9	0.61	0.19	0.38	0			
Geometry and Measurement	7	0.65	0.24	0.31	1	1.22		0.58
Patterns and Algebra	7	0.59	0.17	0.37	1	0.83		0.55
Data Analysis, Probability, and Discrete Mathematics	7	0.56	0.20	0.38	1	1.83		0.60
Problem Solving	12	0.67	0.39	0.42	3	1.29	0.50	0.58

*For open-ended (OE) items, the mean score is the mean of students' scores on a scale of 0 to 3 for mathematics and 0 to 4 for LAL. Writing is scored on a scale of 0 to 5 for grade 5 and 0 to 6 for grades 6 and 7. Note that for grades 6 and 7, the writing scores were doubled in data analyses and score reporting.

Table 4.2.3
Grade 7 - Item Difficulty and Discrimination Summary Statistics for Multiple-Choice and Open-Ended Items by Content Area and Cluster

Test Section/ Cluster	Multiple-choice					Open-ended*				
	Item Difficulty			Item Discrimination		Item Difficulty			Item Discrimination	
	Nitem	Mean	S.D.		Mean	Nitem	Mean	S.D.		Mean
LAL	**20**	**0.68**	**0.15**		**0.39**					
Writing						1	5.84			0.62
Reading	20	0.68	0.15		0.39	4	1.34	0.30		0.52
Interpreting Text	13	0.67	0.14		0.39					
Analyzing Text	7	0.70	0.19		0.38	4	1.34	0.30		0.52
Math	**30**	**0.51**	**0.17**		**0.37**	**3**	**0.74**	**0.11**		**0.57**
Number and Numerical Operation	7	0.55	0.13		0.40	1	0.85			0.58
Geometry and Measurement	9	0.45	0.16		0.35	0				
Patterns and Algebra	7	0.55	0.25		0.37	1	0.74			0.61
Data Analysis, Probability, and Discrete Mathematics	7	0.52	0.14		0.37	1	0.64			0.52
Problem Solving	27	0.52	0.17		0.39	3	0.74	0.11		0.57

*For open-ended (OE) items, the mean score is the mean of students' scores on a scale of 0 to 3 for mathematics and 0 to 4 for LAL. Writing is scored on a scale of 0 to 5 for grade 5 and 0 to 6 for grades 6 and 7. Note that for grades 6 and 7, the writing scores were doubled in data analyses and score reporting.

Table 4.2.4
Grade 5 - Frequency Distributions for Multiple-Choice p-values and Discrimination Indices by Content Area and Cluster

	Nitem	p-value						Discrimination					
		Median	p < 0.25	0.25 <= p < 0.50	0.50 <= p < 0.75	0.75 <= p < 0.90	p >= 0.90	Median	*pb < 0.20	0.20 <= pb < 0.30	0.30 <= pb < 0.40	0.40 <= pb < 0.50	pb >= 0.50
LAL	**20**	**0.77**	**0**	**0**	**6**	**14**	**0**	**0.46**	**1**	**1**	**4**	**14**	**0**
Interpreting Text	12	0.75	0	0	6	6	0	0.45	1	1	3	7	0
Analyzing Text	8	0.80	0	0	0	8	0	0.47	0	0	1	7	0
Math	**30**	**0.70**	**0**	**6**	**12**	**9**	**3**	**0.37**	**2**	**6**	**14**	**8**	**0**
Number and Numerical Operation	7	0.74	0	0	4	3	0	0.40	0	1	3	3	0
Geometry and Measurement	9	0.73	0	2	3	3	1	0.30	1	4	4	0	0
Patterns and Algebra	7	0.66	0	2	2	1	2	0.36	1	0	5	1	0
Data Analysis, Probability, and Discrete Mathematics	7	0.62	0	2	3	2	0	0.41	0	1	2	4	0
Problem Solving	16	0.72	0	5	5	4	2	0.36	1	4	6	5	0

* While ideally, items should have a point-biserial of at least .20, these items had acceptable p-values and were retained to preserve adequate content coverage at the cluster level.

Table 4.2.5
Grade 6 - Frequency Distributions for Multiple-Choice p-values and Discrimination Indices by Content Area and Cluster

	Nitem	p-value						Discrimination					
		Median	p < 0.25	0.25 <= p < 0.50	0.50 <= p < 0.75	0.75 <= p < 0.90	p >= 0.90	Median	*pb < 0.20	0.20 <= pb < 0.30	0.30 <= pb < 0.40	0.40 <= pb < 0.50	pb >= 0.50
LAL	**20**	**0.68**	**0**	**2**	**11**	**7**	**0**	**0.37**	**0**	**6**	**8**	**6**	**0**
Interpreting Text	11	0.59	0	2	8	1	0	0.36	0	5	4	2	0
Analyzing Text	9	0.78	0	0	3	6	0	0.40	0	1	4	4	0
Math	**30**	**0.63**	**1**	**10**	**12**	**6**	**1**	**0.40**	**4**	**5**	**6**	**11**	**4**
Number and Numerical Operation	9	0.66	0	3	4	2	0	0.40	1	1	2	5	0
Geometry and Measurement	7	0.75	0	3	0	3	1	0.29	1	3	0	3	0
Patterns and Algebra	7	0.62	0	2	4	1	0	0.39	1	0	3	2	1
Data Analysis, Probability, and Discrete Mathematics	7	0.63	1	2	4	0	0	0.47	1	1	1	1	3
Problem Solving	12	0.47	0	7	5	0	0	0.43	2	1	2	5	2

* While ideally, items should have a point-biserial of at least .20, these items had acceptable p-values and were retained to preserve adequate content coverage at the cluster level.

Table 4.2.6
Grade 7 - Frequency Distributions for Multiple-Choice p-values and Discrimination Indices by Content Area and Cluster

	Nitem	p-value						Discrimination					
		Median	p < 0.25	0.25 <= p < 0.50	0.50 <= p < 0.75	0.75 <= p < 0.90	p >= 0.90	Median	*pb < 0.20	0.20 <= pb < 0.30	0.30 <= pb < 0.40	0.40 <= pb < 0.50	pb >= 0.50
LAL	**20**	**0.64**	**0**	**2**	**11**	**6**	**1**	**0.40**	**0**	**4**	**6**	**8**	**2**
Interpreting Text	13	0.62	0	1	8	4	0	0.41	0	2	4	6	1
Analyzing Text	7	0.65	0	1	3	2	1	0.38	0	2	2	2	1
Math	**30**	**0.54**	**3**	**11**	**14**	**2**	**0**	**0.40**	**1**	**6**	**9**	**11**	**3**
Number and Numerical Operation	7	0.55	0	3	4	0	0	0.40	0	0	3	4	0
Geometry and Measurement	9	0.47	2	3	4	0	0	0.40	1	2	1	3	2
Patterns and Algebra	7	0.62	1	2	2	2	0	0.38	0	2	2	3	0
Data Analysis, Probability, and Discrete Mathematics	7	0.55	0	3	4	0	0	0.39	0	2	3	1	1
Problem Solving	27	0.50	3	11	12	1	0	0.39	1	6	8	9	3

* While ideally, items should have a point-biserial of at least .20, these items had acceptable p-values and were retained to preserve adequate content coverage at the cluster level.

Table 4.2.7
Number of Dichotomous Items Where at Least One Distractor had a p-value or Point-Biserial Higher than or Equal to that of the Keyed Option

Test	Grade	N-item	p-value*	point-biserial*
LAL	5	20	0	0
	6	20	0	0
	7	20	1	0
Math	5	30	2	1
	6	30	5	5
	7	30	5	2

* The p-value and point-biserial in this table are calculated in the same way as for a correct answer, except in this case the distractor is used instead of the correct answer.

4.3 Speededness

The consequence of time limits on examinee's scores is called speededness. An exam is "speeded" to the degree that those taking it score lower than they would have if the test had not been timed. Most speededness statistics are based on the number of items that were not attempted by students. In each separately timed subsection of a test, if a student does not attempt the last item on the test, it would be assumed that the student may not have reached the item because they ran out of time.

The NJ ASK was not designed to be a speeded test, but rather a power test. That is, all students are expected to have ample time to finish all items and prompts. Because the tests were administered in three days, with multiple sessions each day, students were assumed to have enough time to complete the test. For example, a single writing prompt and a passage with 10 MC and 2 OE items were administered on the first day of testing for all grade levels. It included in two sessions a 25- (grade 5) or 45- (grades 6 and 7) minute session for the writing prompt and a 45-minute session for the reading test. Another passage and 10 MC and 2 OE items were administered on the second day. It included a 45-minute session. The mathematics tests were administered the third day. They included an 89-minute session with a 5-minute silent break after about 45 minutes.

Table 4.3.1 presents the percentage of students omitting the last MC item in each test session. For the LAL test, less than one percent of students omitted the last MC item in each session, while in the mathematics test, less than four percent of students omitted the last MC item.

Table 4.3.1
Percentage of Students Omitting the Last MC Item in Each Test Session

	Session	Item Location	Grade 5	Grade 6	Grade 7
Day One	LAL 2	LAL MC10	1%	1%	1%
Day Two	LAL 3	LAL MC20	0%	0%	0%
Day Three	Math	Math MC 29	0%	1%	0%
		Math MC 30	2%	0%	2%

4.4 Item Bias Statistics

Using data from the 2007 operational test administration, Differential Item Functioning (DIF) was examined with the Mantel-Haenszel (1959)[3] procedure for the MC items and Winsteps (v3.60, Linacre, 2006)[4] for the OE items. Results are summarized in Table 4.4.1. The Mantel-Haenszel (MH) method is a non-parametric approach to DIF. In the MH procedure, total raw scores are held constant while the odds ratio is estimated. In testing practice, the odds ratio is generally converted to the delta metric and the ETS categorization is applied to flag the significance of DIF effects (Dorans & Holland, 1993)[5].

The critical values of the ETS categorizations are 1.00 and 1.50 on the delta scale for categories of A, B, and C. Specifically, if the absolute value of delta is smaller than 1.00 then the item is categorized as "A." If the absolute value of delta is larger than or equal to 1.50, then the item is classified as "C." Otherwise, items are categorized as "B." In both categories of A and C, statistical significance is set at the 5% level for a single item.

DIF detection with Winsteps is a Rasch-model-based approach. According to Linacre, Winsteps DIF detection contains the following steps:

- A joint run with all persons and all items is used to produce anchor values, i.e., ability and rating (or partial credit) scale structure.
- A subgroup run (reference group) with person abilities (or partial credit) scale structure anchored is used to produce group *R* item difficulties (D_R).
- Another subgroup run (focal group) with person abilities (or partial credit) scale structure anchored is used to produce group *F* item difficulties (D_F).
- DIF contrast ($D_F - D_R$) is obtained by taking the different DIF measures of the two-subgroup runs.
- A t-test statistic gives significance values as a unit normal deviate.

Rasch and Mantel-Haenszel procedures for DIF are equivalent under certain conditions (Linacre & Wright, 1989[6]; Schulz, Perlman, Rice, & Wright, 1996[7]). Similar to ETS classifications, the DIF output yielded by Winsteps is classified as negligible (A), slight to moderate (B), or moderate to severe (C). The classification criterion is as follows:

---

[3] Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of National Cancer Institute*, 22, 719-748.
[4] Linacre, M. (2006). Winsteps Rasch Measurement 3.60 [Computer software], Chicago, IL.
[5] Dorans, N. J. & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.
[6] Linacre, J. M. & Wright, B. D. (1989). The equivalence of Rasch, PROX, and Mantel-Haenszel. Transactions of the Rasch Measurement SIG, 3-2, 1-3.
[7] Schulz, E. M., Perlman, C., Rice, W. K., & Wright, B. D. (1996). An empirical comparison of Rasch and Mantel-Haenszel procedures for assessing Differential Item Functioning. In G. Engelhard & M. Wilson (Ed.), Objective measurement: Theory into practice (Vol 3). NJ: Ablex.

If a t-value is smaller than 2.58 or DIF contrast is smaller than 0.45 logits, then the item is flagged as "A." If a t-value is larger than 2.58 and the DIF contrast is larger than 0.65 logits, then the item is flagged as "C." Other items are flagged as "B." This categorization seems to be slightly more conservative than the ETS categorizations using the MH odds ratio (Liu & Mix, 2006)[8].

Table 4.4.1
Frequency Distribution of DIF Categories by Item Type and Grade Level

Test	Grade	Group	Dichotomous*			Open-ended*		
			A**	B**	C**	A**	B**	C**
LAL	5	M/F***	19	0	1	4	1	0
		W/B***	20	0	0	5	0	0
		W/H***	17	3	0	5	0	0
	6	M/F	20	0	0	5	0	0
		W/B	19	1	0	5	0	0
		W/H	20	0	0	5	0	0
	7	M/F	13	4	3	5	0	0
		W/B	19	0	1	5	0	0
		W/H	18	2	0	5	0	0
Math	5	M/F	30	0	0	3	0	0
		W/B	28	2	0	3	0	0
		W/H	28	2	0	3	0	0
	6	M/F	28	2	0	3	0	0
		W/B	30	0	0	3	0	0
		W/H	30	0	0	3	0	0
	7	M/F	30	0	0	3	0	0
		W/B	30	0	0	3	0	0
		W/H	29	1	0	3	0	0

* The Mantel-Haenszel procedure is applied for the MC items and Winsteps for the OE items.
** DIF categories: A, negligible; B, slight to moderate; and C, moderate to severe.
***DIF contrast groups: M/F, Male versus Female; W/B, White versus Black; and W/H, White versus Hispanic.

[8] Liu, Y. & Mix, D. (2006). Categorizations of Winsteps DIF output. Paper presented at the American Educational Research Association conference, San Francisco, CA.

4.5 Summary

The item analyses provided above generally show the NJ ASK to have sound psychometrics properties. For example, the distribution of p-values shows items to be measuring achievement across a broad range of difficulty. Also, item discrimination values show that most items are appropriately correlated with the total test score, and thus contribute to distinguishing between lower performing and higher performing students. In addition to the above, very few students omitted items from the tests. The low percentage of students omitting multiple-choice items provides evidence that the test is a power test of the students' skills, and not their ability to pace themselves through a timed assessment. Finally, item bias statistics based on data from the 2007 operational test administration showed few NJ ASK items to demonstrate DIF. It should be noted that items attaining the third category of DIF (i.e., C) based on field test data are evaluated by an independent sensitivity committee designated by the NJDOE (see Section 2.4 Development of Test Items).

PART 5: TEST ADMINISTRATION

5.1 Introduction

Great care is taken to assure standard administration of the NJ ASK. Close attention to details is necessary to ensure that a student taking the test in one location has an equal opportunity to succeed as a student at another location. Information about the administration of NJ ASK is available in the *New Jersey Assessment of Skills & Knowledge Spring 2007 Test Coordinator Manual Grades 5–7***.** That information will not be fully replicated here, but the following elements are of importance to this technical report.

5.2 Determining Students for Whom a School Is Accountable

State regulations require that all students be included in the statewide assessment program and assessed annually. This includes limited English proficient (LEP) students and students with disabilities. Beginning in school year 2001–2002, students with disabilities were administered the Alternative Proficiency Assessment (APA) for the first time statewide.

All public schools, including those without assessed grades, are counted into the state's accountability system. All schools without assessed grades are counted as one unit with their respective receiving schools. This helps ensure closer vertical alignment of instructional services. In addition, special education students served in proprietary schools are counted in the sending schools' accountability results, which ensure that placement decisions are reviewed closely at the school and district level for optimum student academic performance.

To guarantee compliance with State regulatory requirements (N.J.A.C. 6A:8-4.4), a full academic year is defined as the term that begins on July 1 and ends on or about June 30. This date was established to accommodate the start of the district/school fiscal year and the allowance of academic programs and services offered to students prior to September. Any student enrolling in a school or district for the first time after July 1, up to the test administration date, is not considered to be enrolled for a full academic year. However, for making AYP determinations, a full academic year begins on July 1 and extends to the test administration date.

New Jersey does not include in the accountability system the results of any student enrolled less than one full academic year in a school for school accountability, or in a district for district accountability. This does not exclude from a district's accountability the results of those students who transfer from one school to another within a district.

5.3 Accommodations and Modifications

Testing accommodations and modifications allowed for NJ ASK grades 5–7 are extensive. Special education students must be tested using the modified testing procedures specified in each student's Individual Education Plan (IEP) and must be aligned to the Acceptable Accommodations or Modifications guidelines which are located in Appendix A of the *NJ ASK 2007 TCM Grades 5–7*. Students who are eligible under Section 504 of the Rehabilitation Act of 1973 must be tested using modified testing procedures that are specified in the student's 504 accommodation plan and must be aligned to the Acceptable Accommodations or Modifications guidelines which are located in Appendix A of the *NJ ASK 2007 TCM Grades 5–7*.

Accommodations for Limited English Proficient (LEP) students taking the NJ ASK are also allowed. Limited English Proficient accommodations do not apply to former LEP students, but only to students currently participating in a language assistance program. Accommodations include additional time, translation of the test directions into the student's native language, and use of a bilingual translation dictionary. Electronic dictionaries and Franklin Translators are not permitted. Also, read-aloud directions for LEP students must be modified to reflect the accommodations listed above. It should be noted that LEP students in attendance for less than one academic year are exempt from the LAL assessment, but must take the mathematics assessment and should be enrolled in a language assistance program.

5.4 Testing Exemptions

The April 2007 test administration of the NJ ASK was for all students in grades 5, 6, and 7 in New Jersey. Even students attending out-of-district placements were required to take the NJ ASK. Only out-of-district placements that were approved testing sites could administer the NJ ASK. If a student attended a placement that was not an approved testing site, the student had to return to the home district to be tested or the home district had to send an examiner to the out-of-district placement to administer the test to the student. Test results of students attending out-of-district placements were aggregated into the school and district results of the students' home district. Out-of-district placements included the following:

1. Private schools for the handicapped (in-state or out-of-state)
2. Special services school districts
3. Educational services or joint commissions
4. District-paid placements in regular, alternative, or shared-time public high schools outside of the student's home district (in-state or out-of-state)

5.5 Administration of Tests

Detailed instructions for administering the NJ ASK were provided in the *New Jersey Assessment of Skills & Knowledge Spring 2007 Test Coordinator Manual Grades 5–7***.** The NJ ASK for grades 5–7 was administered according to the following schedule:

- March 12-16; March 19-23, 2007 Regular Testing—LAL – Day 1
- March 12-16; March 19-23, 2007 Regular Testing—LAL – Day 2
- March 12-16; March 19-23, 2007 Regular Testing— Mathematics
- March 26-30, 2007 Makeup Testing

Testing was not to be scheduled immediately after an athletic event or an assembly. All test schedules were checked with the appropriate school officials to ensure that other school activities did not interfere with the test administration. Other test administration procedures included:

- All testing had to be scheduled in the morning. Exceptions included homebound and bedside students, as well as students attending out-of-district placements who were tested at that placement by staff from the student's home district.
- The district and school test coordinators (DTCs/STCs) were responsible for scheduling times and places for regular and make-up testing and for ensuring that all testing was completed according to the procedures and schedule described in the *Test Coordinator Manual* and in the *Examiner Manual*.
- Students who were required to test but were absent for the regular test administration had to be tested on the make-up dates.
- Students whose answer folders were voided during testing were considered to have attempted the test section. They were not allowed to retake or resume taking the voided test section during the make-up.
- Students who began a section of the test and did not complete it during the specified testing time were not allowed to complete the test section during the make-up period or any other time unless additional time was specified in their IEP or 504 plan.

Under the direction of the district test coordinator, the school test coordinator had to prepare a student roster for each examiner. Each roster contained the names of the students whom each examiner would supervise during testing, the names of the proctors assisting the examiner, and the room number. The student rosters were distributed to examiners before testing to allow examiners to prepare for the test administration. Examiners returned the student rosters to the school test coordinator immediately after testing was completed. The school test coordinators kept a copy of each roster and returned the originals to the district test coordinator. The use of student rosters aided test administrators in remaining organized during test administration.

5.6 Test Security Procedures

The NJ ASK test booklets and its contents were treated as secure materials. Detailed procedures for maintaining the security of test materials while they were in the districts were outlined in the *New Jersey Assessment of Skills & Knowledge Spring 2007 Test Coordinator Manual Grades 5–7*. It was the responsibility of the district to guarantee the security of the test materials. Examiners, proctors, and other school personnel were prohibited from copying, reading, discussing, or disclosing any test items before, during, or after test administration. When not being used during a test period, test materials were stored in a secure, locked location that was accessible only to individuals whose access was authorized by the school test coordinator. Inventory forms tracked test materials as they moved from one location to another in districts.

As part of the test development procedures, "breach" test forms and examiner manuals were prepared in the event of a security breach. If the NJDOE identified a security breach during the test administration window, the subcontractor immediately removed the NJ ASK test materials from the involved district or school. The test booklets for the content area affected were coded with a void code indicating a security breach. If the NJDOE determined that there was enough time for testing, the breach forms were delivered to the district and the test was administered to the affected students in the content area impacted by the security breach. For students re-tested during the test administration window, scores were reported based on the breach form. If a security breach was identified after the testing window, the impacted test booklets were coded with a security breach void code and no test results were reported for that content area. However, students received a score for the content area not impacted by the security breach.

5.7 Conclusion

As with any standardized accountability instrument, security is a pressing concern. The information outlined in this chapter and detailed in other cited documents indicates the intense attention paid to the security of the NJ ASK. At every stage detailed attention is devoted to preventing unauthorized access to the questions and to preventing retention by individuals of inappropriate records.

PART 6: SCORING

6.1 Introduction

Acceptable test construction and appropriate administrative procedures would be to no advantage if the scoring of test items were careless, inconsistent, inappropriate to New Jersey's standards, or otherwise ineffective. The following sections are intended to give more detail on the scoring process.

6.2 Multiple-Choice Items

Answer keys approved by the New Jersey Department of Education (NJDOE) were used to score the multiple-choice items after the responses had been scanned. Each item had a specific key (i.e., A, B, C, or D), which had been supplied and verified by the appropriate content specialists. All correct answers were assigned the value of "1" while incorrect answers were assigned the value of "0." At no time in this process was the original scanned answer overwritten, in case the key was determined to be incorrect during the post-scoring quality assurance check. After scoring was completed, simple item statistics were provided to the appropriate RPC content specialist to ensure that the correct keys were being applied. If a key changed, then the process was repeated until the scoring file was correct. The following data were reviewed for quality assurance:

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

6.3 Open-Ended Items and Writing Tasks

The 2007 NJ ASK open-ended questions and writing responses at grades 5, 6, and 7 required hand scoring by Measurement Incorporated (MI) personnel. The processes of selecting and training scorers, reading and scoring papers, and monitoring scoring are described below in detail.

6.4 Scoring Personnel

Because MI has been conducting the handscoring of writing and open-ended items for many years, the company already has available a large pool of qualified, experienced readers. MI informs them that a project is pending and invites them to return. In addition, 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. MI employed many experienced readers for scoring the 2007 NJ ASK and recruited new scorers as well.

MI procedures for selecting new readers are very thorough. After advertising in local newspapers, with a job service, and elsewhere, and receiving applications, staff in MI's human resources department reviews the applications and schedules 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. MI then reviews all the information about each applicant before offering employment.

6.5 Range Finding and Development of Scoring Guides

Range finding meetings were conducted to establish "true" scores for a representative sample of papers. Between 100 and 220 sample papers per task were chosen by MI leadership personnel from the current test administration. For items using specific rubrics, the rubrics were discussed and refined. The sample responses brought to the range finding meetings were 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 consisted of NJDOE content specialists, New Jersey teacher representatives, MI management personnel, as well as the scoring director responsible for each content area and a Riverside content specialist for mathematics.

Each open-ended mathematics item for NJ ASK 5–7 was scored using an item-specific 4-point rubric (score points possible are 0, 1, 2, or 3). Creation of an item-specific rubric for mathematics was accomplished in the following manner:

- All rubrics represented a refinement of the NJDOE generic 4-point rubric which can be found at the following link: http://www.nj.gov/education/njpep/assessment/TestSpecs/MathTestSpec/GEPAMath/HolisticScoringGuide.html
- Each rubric for the mathematics open-ended items (9 rubrics in all, 3 items per grade) was then made item specific, incorporating specific knowledge, skills, and levels of student performance required by the item. This work was performed by Measurement Incorporated.
- The rubrics were further refined in range finding. At range finding, representatives of Measurement Incorporated, Riverside, New Jersey educators, and NJDOE content specialists scored 100 student papers, per item, using the item specific rubrics created by Measurement Incorporated. Further refinement

included incorporating into the rubric specific knowledge, skills, and problem solving strategies and the specific score point they would define.

- Finally, as student papers were scored, scoring questions were referred to the NJDOE content specialist for final adjudication.

After range finding, MI management and the scoring directors developed 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 consisted of three or more annotated examples of each score point in score point order. Training/qualifying sets consisted of clearly anchored papers in random score point order.

6.6 Project Leads

In selecting team leaders, MI's management staff and scoring directors reviewed the files of all the returning staff. Individuals who are experienced team leaders with a record of good performance on previous projects are considered. Also considered are readers who have been recommended for promotion to the team leader position.

It's important to note that MI is an equal opportunity employer that actively recruits minority staff. Historically, temporary staff on major projects average about 70% female, 30% male, 76% Caucasian, and 24% minority. MI does not discriminate 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.

6.7 Training Team Leaders

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 took careful notes on the training papers in preparation for discussion with the scorers. Also, the scoring directors counseled 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 assisted in training scorers in discussions of training sets, and were responsible for distributing, collecting, and accounting for training packets and sample papers during each scoring session. During scoring, team leaders responded to questions, spot-checked scorer packets, and counseled scorers having difficulty with the criteria.

Team leaders also administered the quality control validity sets, monitored the scoring patterns of each scorer throughout the project, conducted retraining as necessary, performed some resolution readings, and maintained a professional working environment. Team leaders worked approximately 7.75 hours per day, excluding breaks.

6.8 Evaluators

Levels of staffing for scoring of the 2007 NJ ASK are listed in Table 6.8.1. The table shows the numbers of scorers, team leaders and scoring directors at each grade level who participated in scoring. Every scorer had a minimum qualification of a 4-year degree, and many had higher degrees.

Table 6.8.1
Number of Scorers, Training Leaders, and Scoring Directors at Each Grade

	2007 NJ ASK		
Grade	Scorers	Team Leaders	Scoring Directors
5	235	25	5
6	240	25	6
7	240	25	5

6.9 Scoring Procedures

All responses requiring hand scoring were scored by MI. All open-ended items for mathematics were scored on a scale of 0 to 3 while LAL was scored on a scale of 0 to 4. Writing was scored on a scale of 0 to 5 for grade 5 and 0 to 6 for grades 6 and 7. Note that for grades 6 and 7, the writing scores were doubled in data analyses and score reporting. Each response was scored at least once. Ten percent of the responses were randomly assigned to a second reader in order to measure inter-reader reliability (see Part 12 of this Technical Report). Scorers scored each student response without seeing any student biographical or demographic information. Evaluators viewed each response and either assigned a score or a condition code.

Training Scorers and Qualifying

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

After contracts and nondisclosure forms were signed, training began. Scorer training followed the same format as team leader training. The scoring director introduced the set

of anchor papers and thoroughly discussed each score point. This presentation was followed by practice scoring on the training sets. Scorers broke into teams to discuss the papers in the training sets. This arrangement gave scorers an opportunity to discuss any possible points of confusion or problems in understanding the criteria in a small group setting.

Team leaders collected the monitor sheets after the scoring of each training set, and recorded results on a customized log which was examined by the scoring director to determine which papers might give scorers difficulty. The scoring director also "floated" from team to team, listening to the team leaders' explanations and adding additional information when necessary. If a particular paper or type of paper seemed to be causing difficulty across teams, the problem was discussed with the room at large to ensure that everyone heard the same explanation.

For qualifying, scorers had to 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 NJDOE was dismissed. All scorers understood this stipulation when they were hired.

Training was carefully orchestrated so that scorers understood how to apply the rubric in scoring the papers, 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, a significant amount of time was allotted for demonstrations of paper flow, explanations of "alerts" and "flagging," and instructions about other procedures which were necessary for the conduct of a smooth project. Scorers generally worked seven hours per day, excluding breaks.

Monitoring for Quality Assurance

MI constantly monitors the quality of each scorer's work throughout every project. Methods used to monitor scorers' scoring habits in scoring NJ ASK included the use of Daily Reader Status Reports.

Each student writing sample was scored holistically by readers using the Registered Holistic Scoring Method. Previously, a different reader from another team read identified packets a second time. At no time did readers see the other's scores. After the scores from each day's work are entered, MI's data application calculated the results and generated a status report. These reports showed the total number of papers read and the percentage agreement of each reader, both perfect and adjacent, for the second-read packets. The reports also showed score point distributions. Scoring directors examined the reports and used the information to determine the need for retraining of individual readers or the group as a whole. It could easily be determined if a reader was consistently scoring "too high" or "too low," as well as the specific score points with which they may be having difficulty. The Daily Reader Status Reports showed not only the current daily totals for each scorer, but also the project-to-date totals.

Retraining was an ongoing process once scoring began. Daily monitoring of completed packets and analysis of agreement rates provided by the Daily Reader Status Reports and

validity packets alerted team leaders and management personnel to individual retraining needs. If it became apparent that a whole team or a whole group was having difficulty with a particular type of response, large group training sessions were conducted. Standard retraining procedures included 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 were dismissed when, in the opinion of the scoring director and the project director, they had been counseled, retrained, and given every reasonable opportunity to improve but were still performing below the acceptable standard.

6.10 Conclusion

Based upon the above procedures, scoring directors and team leaders carefully monitored each scorer's individual performance and the performance of each scoring group as a whole. This allowed any potential scoring drift to be identified and corrected through further training. Scorers were monitored for scoring accuracy and consistency daily. Throughout the scoring period, the reliability of each evaluator was checked through read-behinds and validity packets.

PART 7: SCALING AND EQUATING

7.1 Introduction

This section details the scaling and equating procedures implemented for the 2007 NJ ASK assessments for grades 5-7. Following the scaling and equating plan approved by the NJ DOE, the 2007 NJ ASK operational test scores were equated to the 2006 NJ ASK operational test scores. To accomplish the scaling and equating, the 2007 NJ ASK operational tests were calibrated with Winsteps (Linacre, 2006[9]). Winsteps is designed to produce a single scale by jointly analyzing data resulting from students' responses to both multiple-choice and open-ended items. Multiple-choice items were calibrated using the Rasch model (Rasch, 1960[10], Wright & Stone, 1979[11]; Anderich, 1978[12]), while the partial credit model (Masters, 1982[13]) was used for open-ended items.

Rasch scaling is "a method for obtaining objective, fundamental, linear measures from stochastic observations of ordered category responses" (Linacre, 2006, p.10). In the Rasch model, the probability of a correct response to item *i* given θ is:

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

where θ = latent trait or ability level,
b_i = the difficulty parameter for item *i*.

Similar to other IRT models (Hambleton, 1989[14]; Hambleton & Swaminathan, 1985[15]), the Rasch model requires an assumption of unidimensionality. (Smith, Jr., 2004[16]). Unidimensionality means that all items measure a single construct. If the data fit the model, the measurement units (logits) have the desirable property of maintaining the same size over the whole continuum. These interval measures may then be used in subsequent statistical analyses that assume an interval scale (Smith, Jr., 2004). Also, like other IRT models, the Rasch model allows for separability of parameter estimates (Hambleton, Swaminathan, & Rogers, 1991[17]; van der Linden & Hambleton, 1997[18]).

---

[9] Linacre, J. M. (2006). A *User's Guide to WINSTEPS MINISTEP Rasch-Model Computer Programs*. Chicago

[10] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.

[11] Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: MESA Press.

[12] Anderich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561-573.

[13] Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, *47*, 149-174.

[14] Hambleton, R. K (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). Washington, DC: American Council on Education.

[15] Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory. Principles and applications*. Boston: Kluwer.

[16] Smith, Jr. E. V. (2004). Evidence for the reliability of measures and validity of measure interpretation: A Rasch measurement perspective. In E. V. Smith, Jr. & R. M. Smith, Introduction to Rasch measurement: Theory, models and applications. Maple Grove, MN: JAM Press.

[17] Hambleton, R. K., Swaminathan, H. & Rogers, H. J. (1991). Fundamentals of Items response Theory. Newbury Park, CA: Sage Publications.

That is, the ability estimates of persons are freed from the distributional properties of the specific items attempted. Likewise, the estimated difficulties of items are freed from the distributional properties of specific examinees used in the calibration. This property was useful for the Braille and Large Print test score scaling described below in Section 7.5.

To equate the 2007 NJ ASK operational test to the 2006 test, anchored calibrations were conducted for each content area and grade level. Following the recommendation of the New Jersey Technical Advisory Committee (TAC), the equating was accomplished in two steps. First, the 2007 test scores were equated to the 2006 "base" scale through anchored calibrations. Next, the equated Rasch measures were re-centered to the 2006 "reported" scale. Specific details for the scaling and equating procedures are discussed in Sections 7.3 and 7.4. Section 7.5 focuses on the procedures and results of the equating for the Braille and Large Print tests. Section 7.6 focuses on the procedures and results for several "special equating" analyses requested by the NJDOE. The next section summarizes some properties of the data used for the 2007 scaling and equating.

7.2. Scaling and Equating Data

Sample Size and Distributions

The 2007 NJ ASK scaling and equating data comprised about 25% of the total student population. A total of 141,784 student records were received through a priority processing sample. Among them, 870 (0.61%) student records were invalid and were removed from further analyses. N-counts for the 2007 NJ ASK scaling and equating samples are summarized in Table 7.2.1.

Table 7.2.1
N-Counts for the 2007 NJ ASK Scaling and Equating Samples by Test and Grade

Test	Total	Valid	Invalid
LAL 5	23992	23815	177
LAL 6	25159	24958	201
LAL 7	21741	21471	270
Math 5	23992	23957	35
Math 6	25159	25141	18
Math 7	21741	21572	169
Total	**141784**	**140914**	**870**

[18] van der Linden, W. J. & Hambleton, R. K. (1997). Handbook of modern Item Response Theory. New York: Springer-verlag.

The 25% sample data used for the 2007 scaling and equating was representative of the total student population in terms of District Factor Group (DFG) and other demographic variables, such as gender, ethnicity, economical status, and Current Limited English Proficiency (CLEP). A comparison between data from the 2006 priority districts and the 25% sample data used for equating is presented in Table 7.2.2. Table 7.2.2 shows the difference between the priority districts and the 25% equating sample to be no more than 2.36% for any DFG group across all tests.

Tables 7.2.3 and 7.2.4 present the N-counts for the 2007 scaling and equating samples by DFG, gender, and ethnicity. Note that the sum for males and females will not equal the total in Table 7.2.1 because some students had a missing value for gender. Similarly, some students had a missing value for ethnicity, or marked multiple ethnicities, and therefore the sum for the ethnic groups will not equal the total number of students. Also reported in Tables 7.2.3 and 7.2.4 are the number of economically disadvantaged students as well as CLEP students.

Table 7.2.2
Comparison between the 2007 NJ ASK Equating Sample and the 2006 Priority Districts by DFG and Test

DFG	Priority Districts	LAL 5		LAL 6		LAL 7		Math 5		Math 6		Math 7	
		Obs(%)	Diff	Obs(%)	Diff	Obs(%)	Diff	Obs(%)	Diff	Obs(%)	Diff	Obs(%)	Diff
A	17.58	15.71	-1.86	15.71	-1.86	16.60	-0.98	15.84	-1.74	15.76	-1.81	16.66	-0.92
B	10.51	10.31	-0.20	10.52	0.01	10.42	-0.10	10.40	-0.11	10.68	0.17	10.56	0.05
CD	9.06	8.32	-0.73	9.76	0.70	9.22	0.16	8.27	-0.79	9.75	0.69	9.19	0.13
DE	15.36	13.32	-2.04	13.24	-2.12	13.06	-2.30	13.27	-2.09	13.18	-2.19	13.01	-2.36
FG	12.63	13.13	0.50	13.10	0.48	13.32	0.70	13.09	0.46	13.06	0.43	13.30	0.68
GH	13.46	13.98	0.53	14.01	0.55	13.18	-0.28	13.95	0.49	14.03	0.57	13.11	-0.35
I	18.91	18.50	-0.41	17.58	-1.33	18.60	-0.31	18.44	-0.47	17.47	-1.44	18.58	-0.34
J	1.65	3.65	2.00	3.19	1.55	3.28	1.64	3.64	1.99	3.18	1.53	3.28	1.64
N		1.63	1.63	1.56	1.56	1.30	1.30	1.66	1.66	1.58	1.58	1.30	1.30
R	0.84	1.44	0.60	1.32	0.47	1.01	0.17	1.45	0.60	1.31	0.47	1.01	0.17
Max			**2.00**		**1.56**		**1.64**		**1.99**		**1.58**		**1.64**
Min			**-2.04**		**-2.12**		**-2.30**		**-2.09**		**-2.19**		**-2.36**

Table 7.2.3
Equating Sample N-Counts by Gender and Ethnicity: LAL

Test Grade	DFG	Male	Female	Asian	Black	Hispanic	Indian Alaska	Hawai Pacific	White	EconDis	CLEP
LAL 5	A	1936	1801	61	1133	2058	3	1	461	2811	632
	B	1314	1138	94	492	1016		3	830	1274	144
	CD	976	1004	115	270	192	9	12	1358	524	26
	DE	1618	1550	310	503	416	1	2	1915	683	60
	FG	1559	1564	192	205	186	2	21	2511	259	37
	GH	1709	1617	308	555	326	3	5	2082	418	44
	I	2219	2181	377	313	221		2	3464	222	47
	J	446	422	84	8	16		1	757	3	3
	R	160	183	18	249	47		1	26	207	
	Total	**11937**	**11460**	**1559**	**3728**	**4478**	**18**	**48**	**13404**	**6401**	**993**
LAL 6	A	1969	1947	69	1164	2126	1		533	2877	498
	B	1317	1306	86	576	1074	2	1	846	1323	156
	CD	1263	1170	124	361	222	8	10	1654	641	17
	DE	1713	1588	278	535	382	2	1	2089	709	48
	FG	1683	1583	231	222	203	2	28	2554	279	24
	GH	1774	1718	323	543	310	7	7	2242	431	58
	I	2273	2108	341	304	233	1	3	3482	232	35
	J	406	390	60	8	15		2	700	5	7
	R	164	164	10	231	46			40	222	
	Total	**12562**	**11974**	**1522**	**3944**	**4611**	**23**	**52**	**14140**	**6719**	**843**
LAL 7	A	1835	1723	57	1051	1957	3		467	2526	342
	B	1177	1056	89	385	869		2	854	875	92
	CD	1034	942	105	281	173	2	4	1363	537	13
	DE	1475	1325	226	415	346	1		1794	576	22
	FG	1430	1426	145	184	187	2	16	2315	248	26
	GH	1431	1395	274	460	255	6	2	1783	323	41
	I	2091	1897	318	276	195		3	3185	202	46
	J	354	350	53	12	9	1		625	4	7
	R	103	114	7	163	20			25	164	
	Total	**10930**	**10228**	**1274**	**3227**	**4011**	**15**	**27**	**12411**	**5455**	**589**

Table 7.2.4
Equating Sample N-Counts by Gender and Ethnicity: Mathematics

Test Grade	DFG	Male	Female	Asian	Black	Hispanic	Indian Alaska	Hawai Pacific	White	EconDis	CLEP
Math 5	A	1970	1820	61	1140	2100	3	1	461	2847	680
	B	1331	1158	95	495	1042		3	837	1297	180
	CD	974	1005	116	269	193	9	12	1355	524	33
	DE	1622	1553	315	503	417	1	2	1915	687	67
	FG	1566	1566	192	205	189	2	21	2515	262	45
	GH	1716	1621	315	556	327	3	5	2084	421	55
	I	2224	2189	382	313	224		3	3467	223	62
	J	447	423	84	8	17		1	758	3	4
	R	162	184	18	252	47		1	26	210	
	Total	**12012**	**11519**	**1578**	**3741**	**4556**	**18**	**49**	**13418**	**6474**	**1126**
Math 6	A	1991	1967	71	1171	2155	1		534	2909	539
	B	1345	1336	88	579	1110	2	1	862	1356	213
	CD	1267	1181	126	360	226	9	10	1662	649	28
	DE	1717	1591	280	535	388	2	1	2088	714	56
	FG	1691	1588	234	223	208	3	29	2555	285	34
	GH	1795	1727	336	545	322	7	7	2243	436	84
	I	2276	2111	342	304	235	1	3	3483	232	41
	J	408	391	62	8	15		2	700	5	9
	R	165	164	10	231	47			40	223	1
	Total	**12655**	**12056**	**1549**	**3956**	**4706**	**25**	**53**	**14167**	**6809**	**1005**
Math 7	A	1849	1738	58	1058	1974	3		467	2540	357
	B	1202	1073	91	385	902		2	862	901	132
	CD	1033	946	105	280	176	2	4	1364	539	15
	DE	1475	1326	227	414	349	1		1792	578	24
	FG	1437	1428	145	184	190	2	17	2320	250	29
	GH	1432	1392	275	459	253	6	2	1783	321	40
	I	2100	1901	322	277	198		3	3187	203	57
	J	356	351	54	12	11	1		625	4	10
	R	103	115	7	164	20			25	165	
	Total	**10987**	**10270**	**1284**	**3233**	**4073**	**15**	**28**	**12425**	**5501**	**664**

Descriptive Statistics for the Equating Samples

Table 7.2.5 displays descriptive statistics for raw scores for the equating samples by grade and test content. Table 7.2.6 summarizes descriptive statistics for raw scores for the equating samples by gender. Tables 7.2.7 and 7.2.8 summarize descriptive statistics for raw scores for the samples by DFG. Note that the maximum score is 41 points for Language Arts Literacy (LAL) at grade 5 and 48 points for Language Arts Literacy at grades 6 and 7. The maximum score is 39 points for all the mathematics tests.

Table 7.2.5
Descriptive Statistics for Raw Scores by Grade and Test Content

Test	N	Mean	STD	Min	Max
LAL 5	23815	23.86	5.85	0	36
LAL 6	24958	24.50	6.74	0	43
LAL 7	21471	25.22	6.91	0	43
Math 5	23957	24.48	6.80	0	39
Math 6	25141	22.14	7.57	0	39
Math 7	21572	17.98	7.68	0	39

Table 7.2.6
Descriptive Statistics for Raw Scores by Gender

Test	Gender	N	Mean	STD	Min	Max
LAL 5	Male	12126	23.22	5.95	0	36
LAL 6	Male	12749	23.72	6.76	0	43
LAL 7	Male	11072	24.46	7.13	0	42
Math 5	Male	12206	24.52	6.99	0	39
Math 6	Male	12845	22.09	7.71	0	39
Math 7	Male	11131	18.20	7.97	0	39
LAL 5	Female	11660	24.54	5.66	0	36
LAL 6	Female	12175	25.35	6.58	0	42
LAL 7	Female	10364	26.06	6.56	0	43
Math 5	Female	11722	24.47	6.56	0	39
Math 6	Female	12262	22.21	7.41	0	39
Math 7	Female	10406	17.75	7.35	0	39

Table 7.2.7
Descriptive Statistics for Raw Score by District Factor Group: LAL

Test	DFG	N	Mean	STD	Min	Max
LAL 5	A	3740	20.49	6.44	0	35
	B	2454	22.20	5.82	0	34
	CD	1981	23.75	5.33	0	35
	DE	3173	24.28	5.64	0	36
	FG	3131	24.90	4.86	0	35
	GH	3328	24.71	5.46	0	35
	I	4403	26.02	4.76	0	36
	J	871	27.02	4.26	0	35
	R	343	19.89	5.97	4	33
LAL 6	A	3917	20.52	6.91	0	40
	B	2630	22.11	6.52	0	38
	CD	2436	23.97	6.32	0	42
	DE	3306	25.40	6.25	0	41
	FG	3269	25.51	5.83	0	40
	GH	3504	25.68	6.37	0	41
	I	4381	27.40	5.68	0	43
	J	796	28.66	5.25	0	42
	R	328	19.51	7.08	0	39
LAL 7	A	3562	20.56	7.28	0	40
	B	2297	22.85	6.82	0	40
	CD	1979	24.81	6.41	0	41
	DE	2804	25.94	6.42	0	42
	FG	2864	26.55	5.72	0	41
	GH	2836	26.93	6.22	0	42
	I	3991	28.09	5.74	0	43
	J	704	29.04	4.99	0	41
	R	217	21.05	6.98	0	37

Table 7.2.8
Descriptive Statistics for Raw Scores by District Factor Group: Mathematics

Test	DFG	N	Mean	STD	Min	Max
Math 5	A	3793	21.45	7.07	0	39
	B	2491	22.73	6.55	0	39
	CD	1980	23.83	6.30	0	39
	DE	3180	25.33	6.76	0	39
	FG	3140	25.22	6.10	0	39
	GH	3339	25.19	6.54	0	39
	I	4416	26.55	6.13	0	39
	J	873	27.71	6.17	0	39
	R	346	20.96	7.40	0	39
Math 6	A	3959	18.57	6.97	0	38
	B	2688	19.82	6.94	0	39
	CD	2451	21.06	7.39	0	39
	DE	3313	22.80	7.41	0	39
	FG	3282	23.57	6.95	0	38
	GH	3534	22.87	7.45	0	39
	I	4387	25.08	7.20	0	39
	J	799	26.68	6.46	0	39
	R	329	17.05	7.89	0	38
Math 7	A	3591	13.46	6.30	0	36
	B	2340	15.77	7.05	0	37
	CD	1982	16.70	7.02	0	37
	DE	2805	18.27	7.41	0	39
	FG	2873	18.77	6.90	0	39
	GH	2834	19.79	7.62	0	38
	I	4003	21.51	7.52	0	39
	J	707	22.94	7.13	0	38
	R	218	14.20	7.94	0	37

7.3 Methods for Scaling and Equating

The 2007 NJ ASK test scores were first equated to the 2006 "base" scale and then re-centered to the 2006 "reported" scale. The following steps were implemented to accomplish the scaling and equating:

(1) Re-calibrate the 2006 NJ ASK test scores to create a "base" scale;
(2) Calibrate the 2007 assessment without constraint;
(3) Examine the stability of the common items;
(4) Equate the 2007 assessments to the 2006 re-calibrated "base" scale; and
(5) Re-center the 2007 equated scale to the 2006 original, or "reported" scale.

Raw score to scale score conversion tables are reported in Appendix 7.3. The following sections provide more detail about the procedures and results of the equating for the 2007 NJ ASK.

(1) Re-calibrate the 2006 NJ ASK Test Scores to Create a "Base" Scale

For the 2006 NJ ASK assessments, two scales were differentiated: the original "reported" scale and a subsequent "base" scale. The "base" scale was constructed from more complete data, but only varied slightly from the original "reported" scale for most grades and content areas. The 2006 NJ ASK "reported" scales were created based on calibrations of data from priority districts. The 2006 priority districts data represented about 15% of the 2006 student population and the priority districts were randomly sampled using a stratified sampling procedure. For LAL grades 5 and 6, however, calibrations of the "reported" scales were based on a more complete data set because more students were needed to obtain each score point for some items. Sample sizes of the "reported" calibrations are summarized in Table 7.3.1 by grade and content area.

To accomplish the 2007 NJ ASK equating, the 2006 NJ ASK assessment scores had to be recalibrated to create the "base" scale. For LAL test scores, re-calibrations were based on approximately 100% of the students at each grade. For mathematics, only school districts that were *not* exposed to NJPASS items the previous school year were included. Re-calibrations for mathematics were based on approximately 90% of the total student population at each grade. Sample sizes for the "base" calibrations are summarized in Table 7.3.2 by grade and content area.

The "base" calibrations were implemented with Winsteps and without constraint. The Rasch model fit the data very well. Item mean square ("infit") indices were within the range of 0.70 to 1.30 for all items in each 2006 NJ ASK test. As noted above, item parameter estimates for these recalibrations were regarded as the "base" scale for the 2007 NJ ASK test score equating.

Table 7.3.1
Number of Students Used for Calibrations by Grade and Content Area for NJ ASK 2006 "Reported" Scale

Grade	Number of Students	
	LAL*	Mathematics
5	54,698	14,185
6	49,621	14,208
7	14,182	14,199

* Item calibrations based upon standard setting sample, except LAL for grades 5 and 6 where more student data was needed to obtain each score point.

Table 7.3.2
Number of Students Used for Calibrations by Grade and Content Area for NJ ASK 2006 "Base" Scale

Grade	Number of Students	
	LAL	Mathematics*
5	103,264	91,223
6	103,567	92,777
7	106,874	94,415

* For mathematics, only school districts that were *not* exposed to NJPASS items the previous school year were included.

(2) Calibrate 2007 NJ ASK Assessments without Constraint

The main purpose of this calibration was to examine the stability of common items, or linking items, administered across the two years (i.e., 2006 and 2007). For each test, a calibration was executed "freely" without constraint. For LAL, grade 5, no student obtained the highest number of score points (4) for item 24 (an open-ended item). To keep all the steps for this item as specified in the test blueprint, students obtaining the highest total raw score (36 points, altogether 2 students) were assigned a score of 4 points for the item.

(3) Examine the Stability of Common Items

The stability of common items refers to the expectation that common items function the same way for the groups involved in an equating study. It is recommended that the stability of common items be examined visually and statistically (Kolen and Brennan, 2004[19]). For example, scatter plots can be used to check visually for outlier common items. For NJ ASK, Rasch measures for the common items from the "base" calibrations

[19] Kolen, M. J., & Brennan, R. L. (2004). *Test equating: Methods and practice*. NY: Springer.

and from the 2007 unconstrained or "free" calibrations were plotted against each other. The scatter points for items that function the same should line-up as a straight line. Outlier items will not fall on the straight line and thus can be seen visually.

In addition to visual examination, the stability of common items should be studied analytically. It is recommended that a 0.30-logistic unit be applied as a cut criterion for removing "unstable" common items (Miller, Rotou, & Twing. 2004[20]). That is, any common item that has a difference bigger than 0.30 logits (after adjustments) between the two equating groups should be removed from the common item set and treated as a unique item.

In the 2007 NJ ASK equating study, both visual and analytical methods were applied. Figures 7.3.1 to 7.3.6 present scatter plots by content area and grade that were used for visual examination. Adjusted differences in Rasch logits for anchor items between the 2006 "base" calibrations and the 2007 "free" calibrations are summarized in Tables 7.3.3 and 7.3.4. Note that two items were removed from the common item sets: one item for mathematics at grade 7 (adjusted Rasch difference = 0.3862 logits), and another item for LAL at grade 6 (adjusted Rasch difference = 0.3020 logits).

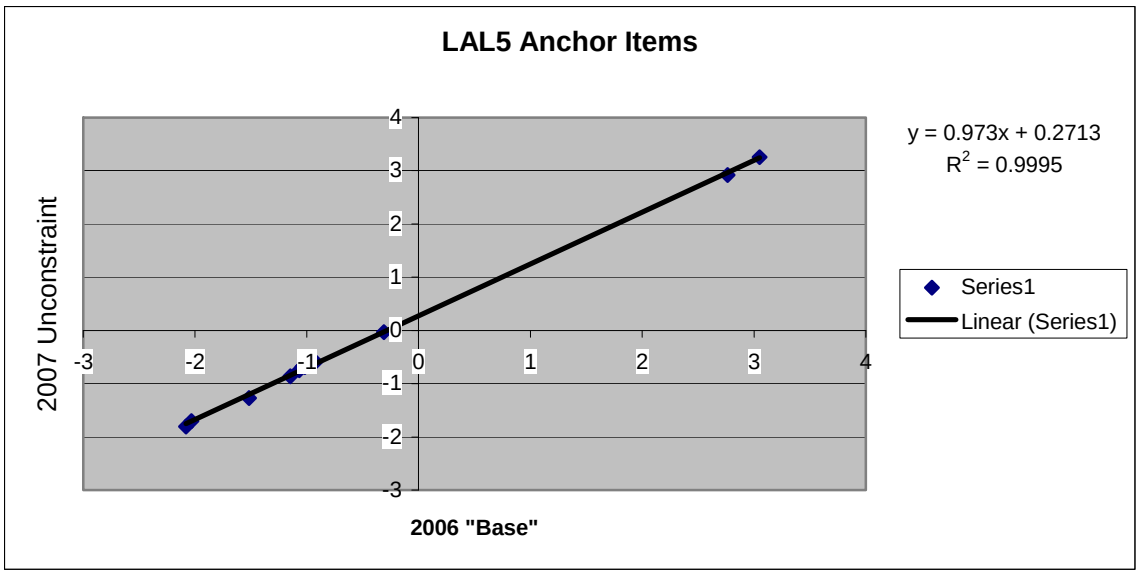

Figure 7.3.1 Scatter Plot for Anchor Items for LAL Grade 5

[20] Miller, G.E., Rotou, O., & Twing, J.S. (2004). Evaluation of the 0.3 logits screening criterion in common item equating. *Journal of Applied Measurement*, 5(2), 172-177.

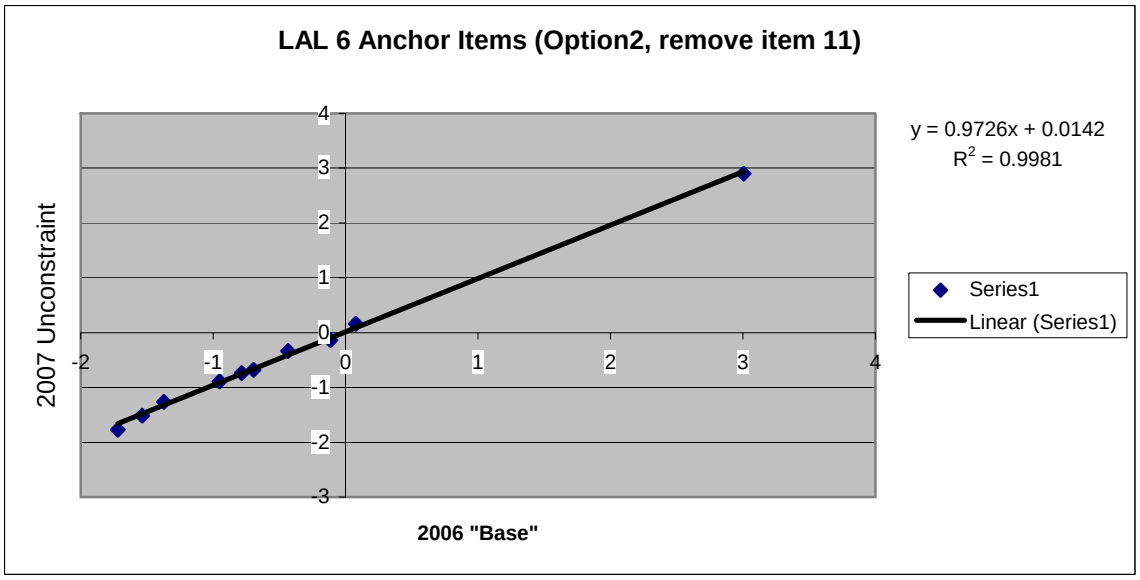

Figure 7.3.2 Scatter Plot for Anchor Items for LAL Grade 6

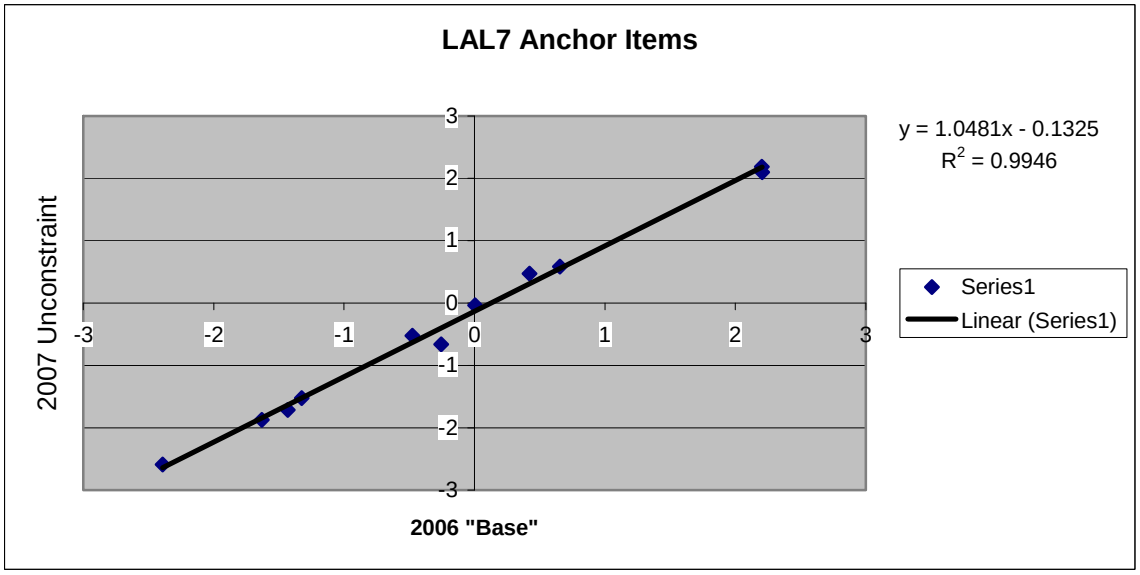

Figure 7.3.3 Scatter Plot for Anchor Items for LAL Grade 7

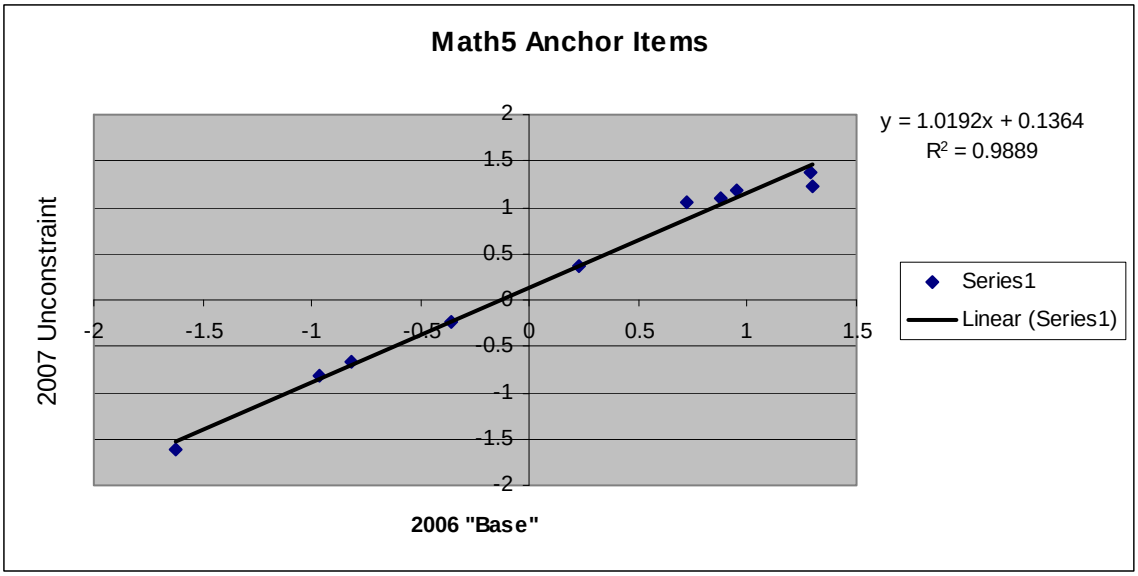

Figure 7.3.4 Scatter Plot for Anchor Items for Mathematics Grade 5

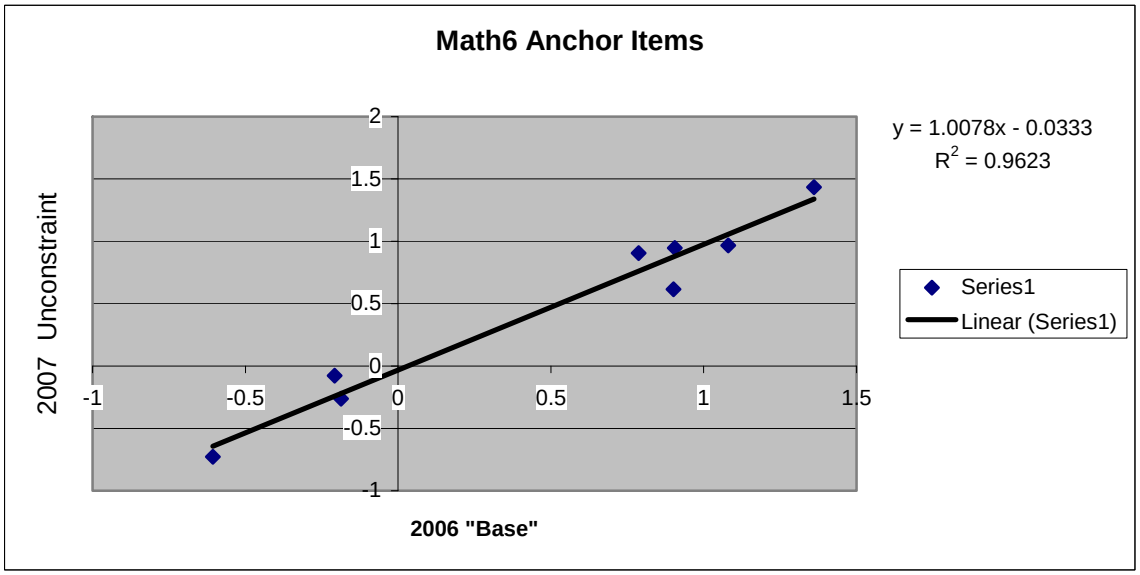

Figure 7.3.5 Scatter Plot for Anchor Items for Mathematics Grade 6

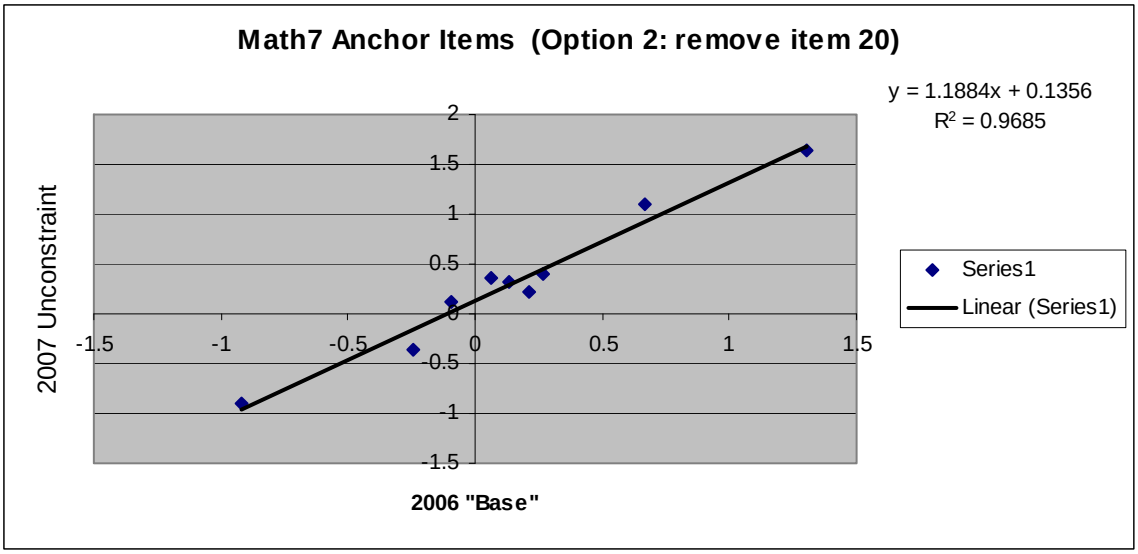

Figure 7.3.6 Scatter Plot for Anchor Items for Mathematics Grade 7

Table 7.3.3
Adjusted Difference in Rasch Logits for Anchor Items between the 2006 "Base" Calibrations and the 2007 "Free" Calibrations (LAL)

LAL 5		LAL 6		LAL 7	
Item 07	Difference	Item 07	Difference	Item 07	Difference
2	-0.03	2	0.04	2	-0.27
3	0.00	3	0.00	3	0.07
4	-0.01	4	-0.01	4	0.10
5	0.00	5	0.07	5	-0.14
6	0.08	6	-0.05	6	-0.10
7	0.05	7	0.06	7	-0.06
8	0.04	8	0.02	8	-0.06
9	0.05	9	-0.07	9	0.10
10	0.05	10	0.08	10	0.19
11	-0.13	12	-0.13	11	0.03
12	-0.08			12	0.12

Table 7.3.4
Adjusted Difference in Rasch Logits for Anchor Items between the 2006 "Base" Calibrations and the 2007 "Free" Calibrations (Mathematics)

Math 5		Math 6		Math 7	
Item 07	Difference	Item 07	Difference	Item 07	Difference
2	-0.01	6	0.15	3	0.12
3	0.08	8	-0.05	7	-0.14
5	-0.01	12	0.10	9	0.04
12	-0.20	19	-0.08	10	0.26
24	0.01	23	0.07	13	0.02
25	-0.13	25	-0.09	25	-0.16
26	0.09	26	-0.26	26	-0.29
28	0.01	30	0.16	28	0.17
31	-0.05			31	-0.03
33	0.20				

(4) Equate the 2007 Assessments to the 2006 Re-calibrated "Base" Scale

It was assumed that the latent trait measured by the 2007 operational tests and the 2006 "base" tests was the same. Given the fact that common anchor items were used across the two years, and that the blueprint and item specifications were the same, it seems reasonable to assume that the underlying latent trait or construct measured by each assessment was the same. To equate the 2007 assessment to the 2006 "base" scale, the Rasch values (difficulties and Rasch-Anderich thresholds for the open-ended items) of the common items were fixed to the 2006 "base" calibrations. This resulted in a raw score to theta conversion on the "base" scale for the 2007 assessment (i.e., the 2007 assessment was scaled on to the 2006 "base" metric).

Tables 7.3.5 and 7.3.6 summarize displacement statistics for the common items generated with the anchored calibrations. A large displacement value indicates lack of convergence, or the presence of anchored or targeted values. It is recommended that "random displacements of less than 0.50 logits are unlikely to have much impact in a test instrument" (Linacre, 2006[21], p. 203). Tables 7.3.5 and 7.3.6 show that all displacement statistics of the common items are smaller than 0.50, indicating the anchored calibrations converged well.

Table 7.3.5
Displacement Statistics for 2007 Anchored Calibrations: LAL

LAL 5		LAL 6		LAL 7	
Item07	Displace	Item07	Displace	Item07	Displace
2	0.02	2	-0.05	2	-0.27
3	0.06	3	-0.09	3	0.07
4	0.04	4	-0.10	4	0.10
5	0.06	5	-0.02	5	-0.13
6	0.13	6	-0.14	6	-0.10
7	0.11	7	-0.03	7	-0.05
8	0.10	8	-0.07	8	-0.05
9	0.10	9	-0.16	9	0.10
10	0.10	10	0.00	10	0.19
11	-0.13	12	0.18	11	-0.05
12	-0.15			12	-0.10

[21] Linacre, J. M. (2006). A user's guide to Winsteps Ministep Rasch-model computer program. Chicago: MESA Press.

Table 7.3.6
Displacement Statistics for 2007 Anchored Calibrations: Mathematics

Math 5		Math 6		Math 7	
Item07	Displace	Item07	Displace	Item07	Displace
2	0.01	6	0.11	3	0.06
3	0.10	8	-0.08	7	-0.21
5	0.00	12	0.06	9	-0.02
12	-0.16	19	-0.12	10	0.21
24	0.03	23	0.02	13	-0.05
25	-0.12	25	-0.13	25	-0.22
26	0.11	26	-0.29	26	-0.35
28	0.03	30	0.12	28	0.09
31	-0.03			31	-0.09
33	0.22				

(5) Re-center the 2007 Equated Scale to the 2006 Original, or "Reported" Scale

A conversion table from the 2006 "reported" scale to the 2006 "base" scale was established so the 2007 equated scale could be re-centered to the 2006 "reported" scale. This was accomplished through number correct raw score because the test is exactly the same for the 2006 "reported" and "base" scales. For example, assume the scales resembled the following:

2006

---Reported---		-----Base-----	
RS	**Theta**	**RS**	**Theta**
1	-2.0	1	-2.5
2	-1.5	2	-2.0
3	-1.0	3	-1.5
4	-0.5	4	-1.0
5	0.0	5	-0.5
6	0.5	6	0.0
7	1.0	7	0.5
8	1.5	8	1.0
9	2.0	9	1.5
10	2.5	10	2.0

Using the above conversion table, a "base" theta of 1.0 (raw score 8) equates to a "reported" theta of 1.5 (also a raw score of 8). While the 2007 assessment was placed on the 2006 "base" scale, raw scores had to be used to re-center the 2007 assessment onto the 2006 "reported" scale. This final step of re-centering the 2007 equated scale to the 2006 "reported" scale was necessary because the 2006 "reported" scale must be maintained over multiple years. The interpolation required to re-center the equated 2007 scale is described in more detail below.

Because the raw scores between 2007 and 2006 do not match as they did between the 2006 "reported" and 2006 "base" scales, interpolation between raw scores and between scale scores on both scales had to be performed to allow raw scores from 2007 to be translated from the "base" scale to the "reported" scale.

The table below shows how this was accomplished:

----------2006----------			-----2007----	
Reported		**Base**	**Base**	
Theta	**RS**	**Theta**	**Theta**	**RS**
-2.0	1	-2.5	-2.3	1
-1.5	2	-2.0	-1.8	2
-1.0	3	-1.5	-1.3	3
-0.5	4	-1.0	-0.8	4
0.0	5	-0.5	-0.3	5
0.5	6	0.0	0.2	6
1.0	7	0.5	0.7	7
1.5	8	1.0	1.2	8
2.0	9	1.5	1.7	9
2.5	10	2.0	2.2	10

In the example above, a 2007 theta of 1.7 is 40% of the way between 1.5 and 2.0 on the 2006 "base" scale. In raw score units, 40% of the way between raw scores 9 and 10 is 9.4. A raw score of 9.4 translates into a "reported" theta of 2.2, which is also 40% of the way between 2.0 and 2.5 on the "reported" scale. A raw score of 9 in 2007 would therefore be equated (or re-centered) to a theta value of 2.2 on the 2006 "reported" scale.

The interpolations were accomplished using an MS Excel calculator that was developed for the purpose of this project and verified through an independent SAS program. Remember that the main task was to link the "base" thetas from the 2007 operational tests to the 2006 "reported" scale.

7.4 Summary of Cut Scores for the 2007 NJ ASK

Total scores for 2007 NJ ASK were reported in scale scores with a range of 100–300. Note that scores of 100 and 300 were a theoretical floor and ceiling and may not actually have been observed for some grades and/or content areas. However, for each test, for a perfect raw score, the scale score was set to 300. A scale score of 200 represents the cut point between Partially Proficient (PP) and Proficient (P) while a scale score of 250 represents the cut point between Proficient and Advanced Proficient (AP). The scale score ranges are as following:

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

To produce the scale score ranges above, linear transformations were applied to theta estimates and scale scores. The following formula was used to obtain the slopes and intercepts for the transformation functions:

$$sc(y) = \left[\frac{sc(y_2) - sc(y_1)}{\theta_2 - \theta_1}\right] y + \left\{ (sc(y_1) - \left[\frac{sc(y_2) - sc(y_1)}{\theta_2 - \theta_1}\right] \theta_1 \right\},$$

where θ_1 and θ_2 are person parameter estimates that correspond to the cut score points, and $sc(y_1)$ and $sc(y_2)$ are scale score points. The above formula was adopted from Kolen and Brennan (2004, p. 337[22]). For NJ ASK in 2006, $sc(y_1)$ was 200 and $sc(y_2)$ was 250. Slopes and intercepts of the transformation functions are summarized in Table 7.4.1. These same slopes and intercepts were applied to the equated, re-centered 2007 theta scale.

Table 7.4.1
Summary of Slopes and Intercepts of Theta to Scale Score Transformation Functions by Grade Level and Content Area

Test	Grade	Proficient			Advanced Proficient			Slope	Intercept
		RS	Theta	SS	RS	Theta	SS		
LAL	5	16	-0.529	200	30	2.266	250	17.8891	209.4633
	6	20	0.025	200	32	1.675	250	30.3030	199.2424
	7	21	0.230	200	34	2.165	250	25.8398	194.0568
Math	5	18	0.099	200	30	1.553	250	34.3879	196.5956
	6	17	-0.096	200	31	1.700	250	27.8396	202.6726
	7	13	-0.579	200	26	0.990	250	31.8674	218.4512

In addition to the above scaling transformation, for the 2007 operational tests, the following rules were applied:

1) The raw score cut (e.g., for Proficient) was selected as the lowest raw score associated with a rounded scale score of 200. The same strategy was also followed for a scale score of 250.
2) If there was no raw score associated with a rounded scale score of 200, the raw score with the highest scale score below 200 was selected as the cut score, and assigned a scale score of 200. For example, if two consecutive raw scores were associated with rounded scale scores of 198 and 201, the scale score of 198 was moved up to 200. The same strategy was also followed for a scale score of 250.
3) Scaled scores below 100 were rounded up to 100.
4) Scaled scores above 300 were rounded down to 300.
5) For each test, for a perfect raw score, the scale score was set to 300.

[22] Kolen, M. J., & Brennan, R. L. (2004). *Test equating: Methods and practice*. NY: Springer.

The raw score cuts for the 2007 NJ ASK and the 2006 NJ ASK "reported" scores are summarized in Table 7.4.2. Note that the difference between the 2007 and 2006 cut scores for each performance level is no more than 1 point. Raw score to scale score conversion tables for the 2007 NJ ASK operation tests are reported in Appendix 7.1.

Table 7.4.2
Raw Score Cuts for the 2007 NJ ASK and the 2006 NJ ASK "Reported" Scores

Scale Score	LAL 5	LAL 6	LAL 7	Math 5	Math 6	Math 7
200	17*(16**)	21(20)	20(21)	18(18)	16(17)	14(13)
250	30(30)	32(32)	33(34)	29(30)	30(31)	27(26)

* 2007 NJ ASK cut scores in raw score metric
**2006 "reported" cut scores

7.5 Methods for Scaling and Equating for Braille and Large Print

Raw score to scale score conversion tables for Braille and Larger Print students are reported in Appendix 7.1. Three steps were followed to create these tables:

1) Re-calibrate the 2006 NJ ASK Braille and Large Print tests to create a "base" scale;
2) Create the 2007 Braille and Larger Print item parameter estimates by recalibrating with the 2007 operational parameter estimates, and;
3) Establish raw score to scale score tables for the Braille and Large Print tests.

Re-calibrate the 2006 NJ ASK Braille and Large Print Tests

For these recalibrations, item parameters, including Rasch measure and Rasch-Anderich step statistics, were fixed to the 2006 "base" calibrations. Items that representatives from the Commission of the Blind deemed inappropriate were not scored. During testing, Braille and Large Print students were instructed to skip these items. The items that were skipped on the Braille and Large Print forms were deleted from the recalibrations. The items that were dropped are summarized in Table 7.5.1.

Table 7.5.1
Items removed for the 2006 Braille and Large Print "base" recalibrations[23]

	LAL 5	LAL 6	LAL 7	Math 5	Math 6	Math 7
Braille	25(Writing)*			1,16	13	32*,33
Large Print				1,16	13	32*,33

* Open-ended item; others are multiple-choice items.

[23] New Jersey Department of Education (2006). Grade 5, 6, and 7 New Jersey assessment of skills and knowledge technical report. P.55.

Create the 2007 Braille and Larger Print Item Parameter Estimates

Similar to the recalibrations described above, item parameters, including Rasch measure and Rasch-Anderich step statistics, were fixed to the 2007 operational test calibrations. Items that were removed from the Braille and Large Print forms were deleted from the calibrations. The items that were dropped are summarized in Table 7.5.2.

Table 7.5.2
Items Removed from the 2007 Braille and Large Print Calibrations

	LAL 5	LAL 6	LAL 7	Math 5	Math 6	Math 7
Braille	14			1	3,32	5,17
Large Print				1		32*

* Open-ended item; all other items are multiple-choice.

Establish Raw Score to Scale Score Tables

Raw score to scale score conversion tables for the Braille and Large Print tests were established using the same procedures described in sections (4) and (5) of Section 7.3. In addition, the same transformations were used as described in Section 7.4.

7.6 "Special Equating" for Students with Lost Responses

A total of 10 "special equating" analyses were conducted for students with lost item responses. For each "special equating" case, scores for one or two items were missing. The missing scores were dropped using the same procedures implemented for the Braille and Large Print tests (Section 7.5). The items that were removed from the "special equating" analyses are summarized in Table 7.6.1. Raw score to scale score conversion tables for each "special equating" are reported in Appendix 7.1.

Table 7.6.1
Items Removed from the 2007 "Special Equating" Calibrations*

	LAL 5	LAL 6	LAL 7	Math 5	Math 6	Math 7
a	11,12		23	12	23,31	23,32
b			24	23,32		10
c			23,24			
d			writing(25)			

* All items are open-ended.

PART 8: REPORTING

8.1 Introduction

Scores are reported in two cycles. Cycle I data are considered preliminary. For additional information about score reports that are very similar to the NJ ASK Cycle I and Cycle II reports, see the GEPA Score Interpretation Manual.[24]

8.2 Cycle I Reports

The Cycle I reports included the following, separate for each grade: Individual Student Report, Student Sticker, Student Rosters by subject, All Sections Roster, Summary of School Performance, and Summary of District Performance. Each Cycle I report is briefly described below.

Individual Student Report

The Individual Student Report (ISR) is a two-sided report, produced in alphabetical sequence for students within the school. The ISR for NJ ASK at grade 5 is shown in sample format as Figure 8.2.1 (front page). Two copies of this report are produced for every student tested, one for the student's permanent folder, and the other for the student's parent/guardian to be shared in a manner determined by the local district.

TEST DATE: MARCH 2007
REPORT PRINTED: 1/30/2007

New Jersey Assessment of Skills and Knowledge
Individual Student Report

NJ ASK

COUNTY: 88 ANY COUNTY
DISTRICT: 7777 ANY DISTRICT
SCHOOL: 666 ANY SCHOOL

Student Information

Student Name: STUDENT A, JAMES
ASK ID No.: 1380282225
District/School ID No.:
Answer Folder No.: 45409 Date of Birth: 06/04/97
LEP: Sex: M
SE: Out-of-District Placement:
Title I: Out-of-Residence Placement:
Grade: 5 Special Form:

Your Child's Scores

Content Area	Scale Score	Proficiency Level
Mathematics	271	ADVANCED PROFICIENT
Language Arts Literacy	270	ADVANCED PROFICIENT

Partially Proficient / Not Pass: Scale Score BELOW 200
Proficient / Pass: Scale Score AT OR ABOVE 200 but BELOW 250
Advanced Proficient / Pass: Scale Score AT OR ABOVE 250

Cluster Scores

Mathematics

The Mathematics section assesses a student's abilities in the following clusters.

Cluster	Your Child's Points	Just Proficient Mean
Number & Numerical Operations	6.0 out of 10.0	4.8
Geometry & Measurement	8.0 out of 9.0	5.1
Patterns & Algebra	8.0 out of 10.0	3.9
Data Analysis, Probability & Discrete Mathematics	8.0 out of 10.0	4.2
Problem Solving	24.5 out of 37.0	16.8
Your Child's Total Points	**30.0 out of 39.0**	

Language Arts Literacy

The Language Arts Literacy section assesses a student's abilities in the following clusters.

Cluster	Your Child's Points	Just Proficient Mean
Writing	4.0 out of 5.0	2.5
Reading	22.0 out of 36.0	13.5
Working with Text	10.0 out of 16.0	5.9
Analyzing Text	12.0 out of 20.0	7.6
Your Child's Total Points	**26.0 out of 41.0**	

88-7777-666
Page 1 of 1

Figure 8.2.1: Individual Student Report

[24] For the GEPA Manual, see: http://www.state.nj.us/njded/assessment/ms/gepa_score_interp_manual.pdf.

The scale scores in LAL and mathematics, along with cluster data, are provided on the front of the ISR (see Figure 8.2.1). Explanatory text about scale scores, proficiency levels, and cluster scores is provided on the back of the ISR.

The Just Proficient Mean is a statewide statistic comprised of the average or mean score attained on each cluster by all students by grade—i.e., general education students (GE), special education students (SE), and limited English proficient students (LEP)—with a scale score of 200—i.e., students who are "just proficient." Students whose NJ ASK test booklets were coded as "void" were excluded from these means, along with any students who took an alternate form (i.e., Braille, Large Print, or Breach).

Student Sticker

The Student Sticker is produced alphabetically, and one sticker for each student within the school is provided. It is a peel-off label designed to be easily attached to the student's permanent record. The scale sores in LAL and Mathematics are provided. Designations of the proficiency levels are printed next to the LAL and mathematics scale scores. Voids, where applicable, are noted.

Student Roster

The Student Roster (Figure 8.2.2) provides a convenient method for reviewing students' test results in either LAL or mathematics. The report displays student names in alphabetical order (last name first).

TEST DATE: MARCH 2007
REPORT PRINTED: 3/12/2007
CYCLE I

New Jersey Assessment of Skills and Knowledge
Student Roster - Mathematics
Grade 5

NJ ASK

COUNTY: 88 ANY COUNTY
DISTRICT: 7777 ANY DISTRICT
SCHOOL: 666 ANY SCHOOL

STUDENTS PROCESSED: 74
(Excludes students who were not scored on the full set of regular items.)

STUDENT NAME / ASK ID NUMBER		DOB	SEX	LEP	SE	TITLE I	504	ACCOM	OUT OF DIST	OUT OF RES	SCALE SCORE	POINTS EARNED BY CLUSTER: NUMBER & NUMERICAL OPERATIONS	GEOMETRY & MEASUREMENT	PATTERNS & ALGEBRA	DATA ANALYSIS, PROBABILITY & DISCRETE MATHEMATICS	PROBLEM SOLVING
												10.0 [1]	9.0	10.0	10.0	37.0
												4.8 [2]	5.1	3.9	4.2	16.8
STUDENT A 1380282225	JAMES	06/04/97	M								271	6.0	8.0	8.0	8.0	24.5
RODRIGUEZ 1380275227	JI YOUNG	10/08/98	F								269	6.0	8.0	8.0	8.0	24.5
STUDENT B 1380289434	REBECCA L	01/11/98	F								266	6.0	8.0	8.0	8.0	24.5
STUDENT X 1395002322	JAMES	02/25/98	M								262	6.0	8.0	8.0	8.0	24.5
PARMA 1380282427	VANESSA	04/11/97	F								253	6.0	8.5	8.0	8.0	24.5
RODRIGUEZ 1380275328	KAITLIN	10/23/98	F								248	6.0	8.0	8.0	8.0	24.5
JACKSON 1380086833	MARC A	05/21/97	M								245	5.0	7.0	8.0	8.0	24.5
STUDENT L 1330027015	JULIO	01/21/95	M								240	5.0	6.5	8.0	8.0	24.5
RATTO 1380195834	MICHELLE M	11/04/98	F								239	5.0	8.0	8.0	9.5	24.5
SCOTT 1380087630	JAVIN	08/09/97	M								236	6.0	7.5	8.0	8.0	21.0

[1] The numbers in this row are the number of possible raw score points.
[2] The numbers in this row are the statewide raw score means for students whose scale score is 200.

88-7777-666

Page 1 of 10

Figure 8.2.2: Student Roster

Users of this report can quickly determine how a particular student performed in each cluster. Following a student's identification information, the student's number of points earned by cluster is printed. State means are also included for reference.

All Sections Roster

The All Sections Roster (Figure 8.2.3) provides a convenient method for reviewing students' complete test results. The report displays student names in alphabetical order (last name first). Users of this report can quickly determine how a particular student performed in both content areas: Language Arts Literacy and mathematics.

TEST DATE: MARCH 2007
REPORT PRINTED: 3/12/2007
CYCLE I

New Jersey Assessment of Skills and Knowledge
All Sections Roster
Grade 5

NJ ASK

COUNTY: 88 ANY COUNTY
DISTRICT: 7777 ANY DISTRICT
SCHOOL: 666 ANY SCHOOL

STUDENTS PROCESSED: 74
(Excludes students who were not scored on the full set of regular items.)

STUDENT NAME / ASK ID NUMBER	DOB	SEX	ETHNIC CODES	LEP	SE	TITLE I	ED	MI	OUT OF DIST	OUT OF RES	TID <1	TIS <1	MATHEMATICS SCALE SCORE	MATHEMATICS PROFICIENCY LEVEL	LANGUAGE ARTS LITERACY SCALE SCORE	LANGUAGE ARTS LITERACY PROFICIENCY LEVEL
JACKSON, MARC A 1380086833	05/21/97	M											245	PROFICIENT	247	PROFICIENT
JONES, MARIUSZ 1380149732	02/06/95	M	W	F1							Y	Y	147	PARTIALLY PROFICIENT	110	PARTIALLY PROFICIENT
JONES, SHERMINE O 1230019722	06/17/96	F	B										150	PARTIALLY PROFICIENT	181	PARTIALLY PROFICIENT
KANG, ROBERT J 1380196128	03/22/97	M											212	PROFICIENT	210	PROFICIENT
KEY, HIRAM 1330028521	08/28/96	M	H										190	PARTIALLY PROFICIENT	220	PROFICIENT
NOVAK, ALEXIS 1380067327	10/25/97	F	BA										188	PARTIALLY PROFICIENT	NOT PRESENT	
OYOLA, ZAKA 1380282528	03/11/98	M	A										167	PARTIALLY PROFICIENT	215	PROFICIENT
PARMA, VANESSA 1380282427	04/11/97	F											253	ADVANCED PROFICIENT	247	PROFICIENT
PATIKOWSKI, DANIEL 1380278634	01/11/96	M	A										203	PROFICIENT	228	PROFICIENT
PAULES, KATHRYN 1380266631	09/26/97	F	H										217	PROFICIENT	239	PROFICIENT

88-7777-666

Page 1 of 10

Figure 8.2.3: All Sections Roster

Following a student's identification information, the student's Scale Score and Proficiency Level (Partially Proficient, Proficient, or Advanced Proficient) are printed for each test section. If the student's test booklet was coded void, the reason code appears in this space.

Performance Summaries (School and District)

Figure 8.2.4 provides an example of the Summary of School Performance report for each content area (Language Arts Literacy and mathematics) and grade. The report is produced at the school level and provides preliminary aggregated data. Final aggregated data are produced in Cycle II reporting as there are inevitably a few changes made to data between Cycle I and Cycle II. Data are provided for total students, general education students, special education students, and limited English proficient students. The report

includes the percent and number of students in each proficiency level for total students, general education students, special education students, and limited English proficient students tested. The report also provides Just Proficient Means as well as scale score means and cluster means for the aforementioned groups.

TEST DATE: MARCH 2007
REPORT PRINTED: 3/26/2007
CYCLE I

New Jersey Assessment of Skills and Knowledge
Summary of School Performance
Grade 7 - Language Arts Literacy

NJ ASK

COUNTY: 88 ANY COUNTY
DISTRICT: 7777 ANY DISTRICT
SCHOOL: 666 ANY SCHOOL

STUDENTS PROCESSED: 22 (Excludes students who were not scored on the full set of regular items.)

PERFORMANCE DATA FOR STUDENTS WITH VALID SCALE SCORES [1]

DEMOGRAPHIC GROUP	Number Processed	APA Students [2]	Not Present	Voids [3]	Valid Scale Scores	Partially Proficient Number	Partially Proficient Percent (%)	Proficient Number	Proficient Percent (%)	Advanced Proficient Number	Advanced Proficient Percent (%)
Total Students [4]	22	1	3	1	17	4	23.5	13	76.5	0	0.0
General Education [5]	15	0	3	1	11	2	18.2	9	81.8	0	0.0
Special Education	2	1	0	0	1	1	100.0	0	0.0	0	0.0
Limited English Proficient [6]	6	0	1	0	5	1	20.0	4	80.0	0	0.0
Current LEP	5	0	0	0	5	1	20.0	4	80.0	0	0.0
Former LEP	1	0	1	0	0	0	0.0	0	0.0	0	0.0
Students Coded SE & Current LEP:	0	0	0	0	0						

SCALE SCORE MEANS & CLUSTER MEANS FOR STUDENTS WITH VALID SCALE SCORES

	Scale Score Mean [7]	Writing	Reading	Working with Text	Analyzing Text
Total Points Possible		12.0	36.0	12.0	24.0
Just Proficient Mean [8]		4.5	16.5	7.0	9.5
Total Students [4]	210.3	4.3	15.7	6.9	8.8
General Education [5]	214.3	4.2	16.1	6.8	9.3
Special Education	162.0	4.0	5.5	2.0	3.5
Limited English Proficient [6]	211.2	4.6	16.9	8.0	8.9
Current LEP	211.2	4.0	16.9	8.0	8.9
Former LEP	0.0	0.0	0.0	0.0	0.0

[1] Percentages may not total 100 due to rounding.
[2] Students who are required to take the Alternate Proficiency Assessment (APA) instead of NJ ASK.
[3] Includes students coded LEP Exempt (LAL Only).
[4] Students are included in Total Students only once, but they appear in all other categories that apply.
[5] Includes students coded Former LEP who are not Special Education.
[6] Includes students coded Current and Former LEP.
[7] The range of scale scores is 100 to 300.
[8] Just Proficient Mean is the statewide raw score mean for students whose scale score is 200.

88-7777-666

Page 1 of 1

Figure 8.2.4: Summary of School Performance

There is also a Summary of District Performance report for each content area and grade, which provides aggregated data for the district. This report includes data for total students, general education students, special education students, and limited English proficient students. The report format is the same as the Summary of School Performance.

8.3 Cycle II Reports

The Cycle II reports, produced separately for each grade, include the following: Cluster Means for Students with Valid Scale Scores (one for each subject), Performance by Demographic Group School, Performance by Demographic Group District, Performance by Demographic Group DFG, and Performance by Demographic Group Statewide. These Cycle II reports are briefly described below.

Cluster Means for Students with Valid Scores

Figure 8.3.1 provides an example of the Cluster Means for Students with Valid Scores report which is created for each content area (Language Arts Literacy and mathematics) and grade. The mean number of raw score points is reported for several important student groups and is aggregated at school, district, DFG group and the state level.

TEST DATE: MARCH 2007
REPORT PRINTED: 3/14/2007
CYCLE II

New Jersey Assessment of Skills and Knowledge
Cluster Means for Students with Valid Scale Scores[1]
Grade 6 - Mathematics

NJ ASK

COUNTY: 88 ANY COUNTY
DISTRICT: 7777 ANY DISTRICT
SCHOOL: 666 ANY SCHOOL

	SCHOOL MEAN	DISTRICT MEAN	DFG A MEAN	STATE MEAN	TOTAL POINTS POSSIBLE	JUST PROFICIENT MEAN [2]
Number & Numerical Operations					9	4.7
Total Students [3]	2.1	2.1	2.1	2.1		
General Education [4]	2.0	2.0	2.0	2.0		
Special Education	0.0	0.0	0.0	0.0		
Limited English Proficient [5]	2.2	2.2	2.2	2.2		
Current LEP	2.3	2.3	2.3	2.3		
Former LEP	2.0	2.0	2.0	2.0		
Geometry & Measurement					10	4.8
Total Students [3]	3.1	3.1	3.1	3.1		
General Education [4]	2.9	2.9	2.9	2.9		
Special Education	0.0	0.0	0.0	0.0		
Limited English Proficient [5]	3.8	3.8	3.8	3.8		
Current LEP	3.6	3.6	3.6	3.6		
Former LEP	4.5	4.5	4.5	4.5		
Patterns & Algebra					10	3.1
Total Students [3]	4.9	4.9	4.9	4.9		
General Education [4]	4.7	4.7	4.7	4.7		
Special Education	0.0	0.0	0.0	0.0		
Limited English Proficient [5]	5.1	5.1	5.1	5.1		
Current LEP	5.4	5.4	5.4	5.4		
Former LEP	4.0	4.0	4.0	4.0		
Data Analysis, Probability & Discrete Mathematics					10	4.4
Total Students [3]	4.9	4.9	4.9	4.9		
General Education [4]	4.5	4.5	4.5	4.5		
Special Education	0.0	0.0	0.0	0.0		
Limited English Proficient [5]	6.0	6.0	6.0	6.0		
Current LEP	5.8	5.8	5.8	5.8		
Former LEP	7.0	7.0	7.0	7.0		
Problem Solving					37	15.9
Total Students [3]	10.3	10.3	10.3	10.3		
General Education [4]	9.2	9.2	9.2	9.2		
Special Education	0.0	0.0	0.0	0.0		
Limited English Proficient [5]	12.9	12.9	12.9	12.9		
Current LEP	13.1	13.1	13.1	13.1		
Former LEP	12.0	12.0	12.0	12.0		

[1] Excludes students coded Void, APA, Not Present, or students who were not scored on the full set of regular items.
[2] The numbers in this column are the statewide raw score means for students whose scale score is 200.
[3] Students are included in Total Students only once, but they appear in all other categories that apply.
[4] Includes students coded Former LEP who are not Special Education.
[5] Includes students coded Current and Former LEP.

88-7777-666

Page 1 of 1

Figure 8.3.1: Cluster Means for Students with Valid Scores

The school, district, DFG, and statewide reports provide a complete analysis of student performance for each content area tested. These reports summarize student performance on the total test sections, disaggregated by population, gender, ethnicity, economic status and migrant status. An example of the district report is given in Figure 8.3.2.

TEST DATE: APRIL 2006
REPORT PRINTED: 03/26/2007
CYCLE II

New Jersey Assessment of Skills and Knowledge
Performance by Demographic Group
District - Grade 6

COUNTY: 88 ANY COUNTY
DISTRICT: 7777 ANY DISTRICT

Demographic Group	Mathematics									Language Arts Literacy								
						Performance Data for Students with Valid Scale Scores [2]									Performance Data for Students with Valid Scale Scores [2]			
	Number Enrolled	APA Students [1]	Number Not Present	Number of Voids	Number of Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	Number Enrolled	APA Students [1]	Number Not Present	Number of Voids [3]	Number of Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean
Total Students [4]	35	0	5	1	29	65.5	34.5	0.0	187.0	35	1	4	1	29	24.1	75.9	0.0	208.2
General Education [5]	26	0	1	1	24	62.5	37.5	0.0	188.4	26	0	4	1	21	9.5	90.5	0.0	216.5
Special Education	4	0	4	0	0	0.0	0.0	0.0	0.0	4	1	0	0	3	66.7	33.3	0.0	174.7
Limited English Proficient [6]	6	0	0	0	6	83.3	16.7	0.0	181.2	6	0	0	0	6	50.0	50.0	0.0	199.2
Current LEP	5	0	0	0	5	80.0	20.0	0.0	180.2	5	0	0	0	5	60.0	40.0	0.0	193.4
Former LEP	1	0	0	0	1	100.0	0.0	0.0	186.0	1	0	0	0	1	0.0	100.0	0.0	228.0
Gender [7]																		
Female	13	0	2	0	11	81.8	18.2	0.0	179.4	13	0	1	0	12	41.7	58.3	0.0	196.2
Male	22	0	3	1	18	55.6	44.4	0.0	191.6	22	1	3	1	17	11.8	88.2	0.0	216.7
Migrant Status																		
Migrant	2	0	1	0	1	100.0	0.0	0.0	167.0	2	1	0	0	1	100.0	0.0	0.0	172.0
Non-Migrant	33	0	4	1	28	64.3	35.7	0.0	187.7	33	0	4	1	28	21.4	78.6	0.0	209.5
Ethnicity [8]																		
White	8	0	1	0	7	14.3	85.7	0.0	204.1	8	0	1	0	7	0.0	100.0	0.0	226.4
Black or African American	6	0	0	1	5	100.0	0.0	0.0	178.2	6	0	1	0	5	0.0	100.0	0.0	216.0
Asian	1	0	0	0	1	100.0	0.0	0.0	180.0	1	0	0	0	1	100.0	0.0	0.0	187.0
Pacific Islander	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0	0	0.0	0.0	0.0	0.0
Hispanic or Latino	15	0	2	0	13	76.9	23.1	0.0	180.8	15	0	2	1	12	41.7	58.3	0.0	192.8
Amer. Indian/AK Native	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0	0	0.0	0.0	0.0	0.0
Other	5	0	2	0	3	66.7	33.3	0.0	190.3	5	1	0	0	4	25.0	75.0	0.0	218.0
Economic Status																		
Econ. Disadvantaged	2	0	0	0	2	100.0	0.0	0.0	163.0	2	0	0	0	2	0.0	100.0	0.0	203.5
Non-Econ. Disadvantaged	33	0	5	1	27	63.0	37.0	0.0	188.7	33	1	4	1	27	25.9	74.1	0.0	208.6

1 Students who are required to take the Alternate Proficiency Assessment (APA) instead of NJ ASK.
2 Percentages may not total 100 due to rounding.
3 Includes students coded LEP Exempt (LAL Only).
4 Students are included in Total Students only once, but they appear in all other categories that apply.
5 Includes students coded Former LEP who are not Special Education.
6 Includes students coded Current and Former LEP.
7 Excludes students who did not have Gender coded.
8 Students who did not have any Ethnicity coded and students with multiple Ethnicities coded are reported in the Other category only.

88-7777

Page 1 of 1

Figure 8.3.2: Performance by Demographic Groups

8.4 State Summary Reporting

The statewide student data file summarizes student performance by grade and content areas tested on the total test sections, disaggregated by population, gender, ethnicity, economic status and migrant status. This data file is made available via the NJDOE's Office of Evaluation and Assessment's Web site.

8.5 Interpreting Reports

The NJ ASK score report information is used for the purpose of district monitoring. The data are also provided to assist districts in the review of current curricular programs. With the adoption of the New Jersey Core Curriculum Content Standards in May 1996, all districts were required to implement standards based instruction. NJ ASK results displayed in school-level and district-level reports can provide meaningful information for educational program reviews.

All other factors being equal, the reliability (stability) of scores decreases as the number of items used decreases. Generally speaking, reliability is lower in clusters that have smaller numbers of items. All else being equal, differences in mean cluster scores for clusters with smaller numbers of items must be greater than differences for clusters with large numbers of items before they can be considered meaningful. Decreases in reliability also increase the need for multiple measures, particularly where the number of students in the assessed group is small.

All clusters cannot be assumed to be of equal difficulty level. Cluster scores should, therefore, be compared to their respective Just Proficient Means to facilitate effective interpretation. Insofar as tests are not equated at the cluster level, cluster scores cannot be compared from year to year. Year-to-year comparisons should be limited to total test scores in the content areas tested. For each content area, it is the whole test level (only) for which scores are equated.

The NJ ASK reports provide information on clusters in content areas that need further attention. However, since some clusters were assessed with a relatively small number of items, evaluation of a student's performance should never be based solely on the results of the NJ ASK or any other single form of formal or informal assessment. Insofar as the NJ ASK is equated at the test level only, cluster performance should not be directly compared across multiple test administrations.

PART 9: ACCOUNTABILITY

9.1 Introduction

The 2001 re-authorization of the Elementary and Secondary Education Act of 1965 was signed into federal law January 8, 2002. Characterized in the statute as "An Act to close the achievement gap with accountability, flexibility, and choice, so that no child is left behind," it carries the short title of the No Child Left Behind (NCLB) Act of 2001. Like New Jersey, many states have modified and/or supplemented their student assessments to comply with the federal statute and now use assessment results to make both federal and state accountability decisions. This section of the Technical Report focuses primarily on accountability related to NCLB.

9.2 Accountability Model - Overview

Adequate Yearly Progress (AYP) is the term used in NCLB to refer to the minimum improvement required of each school and district over the course of one year. It is measured at the school and district levels by:

- Measuring growth in the percentage of students scoring Proficient or above in reading and mathematics.
- Assessing improvement on one "other academic indicator."
- Testing at least 95% of enrolled students and student subpopulations of sufficient size.

As the term AYP suggests, progress toward NCLB academic goals is evaluated annually. New Jersey's definition of AYP is determined by a formula. The formula calculates the number of proficient scores over the number of valid test scores, with 20% of the items responded to denoting a valid test score. Standards have been set based on starting points and incremental increases aimed at 100 percent proficiency by 2014. Separate starting points for this process have been set for LAL and mathematics for grades 4, 8, and 11.

9.3 Accountability Model - Goals

Using the Compound Annual Growth Rate (CAGR) model, state benchmarks for the standards will be raised every three years in school years 2004–2005, 2007–2008, 2010–2011, and 2013–2014 until the 100 percent proficiency goal is reached for all subject areas at all levels. This methodology employs equal increments of growth calculated on a percentage rate for closing the achievement gap, rather than a straight numerical calculation. Table 9.3.1 presents the annual and intermediate goals based upon application of the CAGR method.

Table 9.3.1
Annual and Intermediate Goals Based Upon Application of CAGR Method

Content Area	Grade	2002-2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2013-2014
LAL	4	68	68	75	75	75	82	82	82	91	91	91	100
	8	58	58	66	66	66	76	76	76	87	87	87	100
	12	73	73	79	79	79	85	85	85	92	92	92	100
Math	4	53	53	62	62	62	73	73	73	85	85	85	100
	8	39	39	49	49	49	62	62	62	79	79	79	100
	12	55	55	64	64	64	74	74	74	86	86	86	100

9.4 Accountability Classification Results

Preliminary AYP status under NCLB accountability requirements for 2007 can be found at the following address:

http://www.nj.gov/njded/title1/accountability/profiles/07/

PART 10: QUALITY CONTROL PROCEDURES

10.1 Quality Control in Data Preparation

Measurement Incorporated (MI) developed a set of rules that data correction staff used to hand-edit information contained on a response document. Data correction provided an opportunity to review and correct the information collected by scanning, ensure that it was scanned accurately, and verify that the information provided by the schools was consistent and correct.

To ensure accuracy, once a correction had been made to the database, the document was validated again to ensure the corrected edit did not create another error. All edits were recorded and tracked in MI's databases, along with the user ID of the staff member making the edits. Performance was constantly and consistently reviewed to ensure that accuracy was maintained and to correct any detected problems.

10.2 Quality Control in Scanning

Scanning and scoring programs were fully tested and reviewed using structured testing methodologies before live test materials were processed, and were continually monitored throughout the process. MI's Quality Assurance (QA) staff developed independent queries to validate all software programs and programmatically produced deliverables for reporting. Each program was tested to ensure that data were included or excluded as appropriate (with particular attention to any special equating situations), and programmatic calculations were performed accurately and according to the reporting rules provided by Riverside and the New Jersey client. During the QA process, reader score sheets were reviewed and compared to student records to verify that scores were applied appropriately. A selection of students was presented to ensure coverage of each type of demographic coding scenario as well as any overrides that were done by MI according to coding rules developed in conjunction with Riverside and the New Jersey client.

MI monitored all aspects of machine scanning. Ensuring the accuracy of demographic data collection was an important component of producing accurate student score reports. Therefore, MI created a detailed data verification plan according to their usual high standards for data capture. This plan encompassed all phases and was a comprehensive set of quality processes to ensure the utmost accuracy of the final reports and file deliverables.

QA staff conducted rigorous tests prior to the scanning of live answer documents to collect student demographic data. Scanning applications that included every scannable document were written using Pearson's ScanTools Plus® application. Each application was tested to ensure it was properly defined and set up. This testing stage was conducted to ensure that the data derived from all grids appearing on the scannable document were included in the export file, were accurately read, and returned the correct value. A quality control sample of answer document demographics (test deck) was created so that all possible responses were verified. This structured method of testing provided exact test

parameters and a methodical way of determining that the output received from the scanner(s) was correct. The documents and the data file created from them were carefully compared to further ensure that results from the scanner were accurate. Accurate scanner calibration was verified at the time of testing, and scanners were re-calibrated to specifications prior to each staff shift change to ensure that calibration remained constant and accurate.

MI has developed a set of comprehensive guidelines for eliminating situations that might threaten the integrity of scanned data. By following these strict guidelines, our scanner operators ensured that the most accurate information possible was read from the document. Scanner operators handled minor response document repairs that allowed the original documents to go through the scanner properly. Small rips in a page were often repaired using cellophane tape, for example. In the rare event that a page from an answer document had more serious damage, the gridded responses from the original, damaged page were transcribed onto a replacement page. A second person verified that the page was transcribed correctly. An adhesive label was placed on the original page explaining that it was transcribed, who transcribed it, and the litho code value (answer folder number) of the page it was transcribed onto. This page was kept with the rest of the document as a reference in case of a question or challenge.

Besides handling student document pages that do not scan, scanner operators also responded to extra pages rejected by the scanner. When an extra page contained a handwritten or typewritten response, the scanner operator filled out a label identifying the document it was associated with and attached that label to the page. The scan bin was set aside, and a scoring assistant was notified. The scoring assistant determined whether the page contained responses that should be used in determining the student's score. If it did, the item with which the extra page was associated was indicated on the label. This extra page was kept with the corresponding original response document page throughout processing so that scoring staff would assign the correct score to the student.

10.3 Quality Control in Editing and Data Input

MI used a successive check of quality assurance and control system to ensure and maintain accurate and timely scoring results, reporting, and dissemination of data. Throughout the execution of the software testing, all defects were logged, assigned, and followed through to resolution. Software changes or "fixes" provided by the developer to resolve defects were re-tested until satisfactory results were achieved. Regression testing of previously tested functionality was performed to ensure that the fix did not adversely affect any other functionality of the application/system.

Deployment of software applications to the staging environment was also tested during the QA process in partnership with MI's Network Operations (NetOps) team. The staging environment closely matched the production environment, which enabled us to determine projected behavior once the application was deployed to the production environment.

10.4 Quality Control in Scoring

MI constantly monitors the quality of each scorer's work throughout every project. Methods used to monitor scorers' scoring habits in scoring NJ ASK included the use of Daily Reader Status Reports.

For writing and open-ended items, each student writing sample was scored holistically by readers using the Registered Holistic Scoring Method. Previously, a different reader from another team read identified packets a second time. At no time did readers see the other's scores. After the scores from each day's work were entered, MI's data application calculated the results and generated a status report. These reports showed the total number of papers read and the percentage agreement of each reader, both perfect and adjacent, for the second-read packets. The reports also showed score point distributions. Scoring directors examined the reports and used the information to determine the need for retraining of individual readers or the group as a whole. It could easily be determined if a reader was consistently scoring "too high" or "too low," as well as the specific score points with which they may have been having difficulty. The Daily Reader Status Reports showed not only the current daily totals for each scorer, but also the project-to-date totals.

Retraining was an ongoing process once scoring began. Daily monitoring of completed packets and analysis of agreement rates provided by the Daily Reader Status Reports and validity packets alerted team leaders and management personnel to individual retraining needs. If it became apparent that a whole team or a whole group was having difficulty with a particular type of response, large group training sessions were conducted. Standard retraining procedures included 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 were dismissed when, in the opinion of the scoring director and the project director, they had been counseled, retrained, and given every reasonable opportunity to improve but were still performing below the acceptable standard.

10.5 Quality Control in Reporting

MI fully recognizes the importance of problem-free score reporting and has employed stringent quality control procedures ensuring that reporting on all levels was complete and accurate to the extent possible for the NJ ASK 5–7 assessment. With this in mind, MI thoroughly tested, reviewed, and proofread all reporting deliverables prior to delivery to Riverside and the New Jersey client.

QA staff verified the content of preliminary reports during the preliminary reporting phase and ensured that reports contained the correct information presented in a clear, concise manner. Reports were tested to ensure that valid values were verified, valid codes were included on student records, correct scores were reflected and were attributed to the correct student, cluster scores were accurately aggregated and totaled, and appropriate student totals were reported in all aggregate reports.

QA also verified formatting of reports, as designed by Riverside and the New Jersey client, including fonts, footnotes, line separations, sections, and headings. This testing process was included in all aspects of data files, electronic reports, and printed reports. During the printing of the final reports, QA verified that print quality was excellent and all reports for all students, schools, and school systems were complete.

10.6 Additional Quality Control Procedures

Listed below are QC steps Riverside Publishing used in addition to those discussed above:

- Compared item p-values to definition keys.
- Construct keys from item data by recoding all but the correct option (i.e, A, B, C, D) to missing and then obtaining the frequency distribution. Compared this with the key definition keys.
- Obtained frequency distribution of every demographic variable and score (item, cluster, and total).
- Re-scored all items and compared to item file score vector (or file total score).
- Re-scored all items and clusters and compared to file cluster score.
- Replicate all RS to SS merges and checked the resulting distributions.

PART 11: SUMMARY STATISTICS

11.1 Introduction

This section provides descriptive statistics for number correct raw score and for scale scores. Statistics include N-counts, means, standard deviations, minimum and maximum values, and a variety of data disaggregations, including student demographic group and District Factor Group (DFG).

11.2 Descriptive Statistics for Total Raw Score

Descriptive statistics of total scores for NJ ASK 2007 are summarized in Table 11.2.1 by test content, form, and grade level. A total of 309,203 students participated in the LAL grades 5–7 tests, and 310,963 students participated in the mathematics grades 5–7 tests.

Table 11.2.1
Descriptive Statistics for Total Raw Scores by Content Area and Grade Level

Test	Form	Grade	N	Mean	STD	Min	Max	Nitem	Max Possible
LAL	OP	5	101265	23.92	5.63	0	38	25	41
	OP	6	103146	24.53	6.53	0	45	25	48
	OP	7	104350	25.05	6.79	0	46	25	48
	AL	6	118	22.76	5.51	5	35	25	48
	AL	7	8	12.38	5.88	2	23	25	48
	BR	5	6	20.00	6.54	10	27	24	40
	BR	6	9	26.56	6.80	14	35	25	48
	BR	7	3	16.00	12.29	7	30	25	48
	LP	5	99	19.24	6.75	4	31	25	41
	LP	6	103	19.31	6.75	3	35	25	48
	LP	7	96	20.24	7.77	2	37	25	48
Math	OP	5	101941	24.38	6.60	2	39	33	39
	OP	6	103689	22.17	7.45	0	39	33	39
	OP	7	104902	17.80	7.66	0	39	33	39
	AL	6	116	21.91	6.22	7	34	33	39
	AL	7	8	6.88	2.23	4	11	33	39
	BR	5	6	19.50	6.63	8	28	32	38
	BR	6	9	20.11	7.96	8	30	31	37
	BR	7	3	10.67	9.81	5	22	31	37
	LP	5	97	19.95	7.98	3	35	32	38
	LP	6	100	16.60	7.46	1	37	33	39
	LP	7	92	13.86	7.14	3	34	32	36

*OP: Operational Test; AL: Alternative Operational Test; BR: Braille; and LP: Large Print.

11.3 Descriptive Statistics for Total Raw Scores by Cluster

Tables 11.3.1 through 11.3.3 summarize the means and standard deviations for number correct raw score by cluster for the 2007 NJ ASK operational test forms.

Table 11.3.1
Grade 5 - Means and Standard Deviations for Number Correct Raw Score for the 2007 NJ ASK Operational Forms

	Number of Items		Number of Possible Points	Raw Score		Average Percent Correct
	MC	OE		Mean	Standard Deviation	
LAL	**20**	**5**	**41**	**23.92**	**5.63**	**58.34%**
Writing		1	5	2.78	0.79	55.60%
Reading	20	4	36	21.14	5.19	58.72%
Working with Text	12	1	16	10.73	2.82	67.06%
Analyzing Text	8	3	20	10.41	2.73	52.05%
Math	**30**	**3**	**39**	**24.38**	**6.60**	**62.51%**
Number and Numerical Operation	7	1	10	6.72	2.33	67.20%
Geometry and Measurement	9	0	9	6.27	1.56	69.67%
Patterns and Algebra	7	1	10	6.14	1.97	61.40%
Data Analysis, Probability, and Discrete Mathematics	7	1	10	5.25	2.30	52.50%
Problem Solving	16	3	25	14.22	4.50	56.88%

Table 11.3.2
Grade 6 - Means and Standard Deviations for Number Correct Raw Score for the 2007 NJ ASK Operational Forms

	Number of Items		Number of Possible Points	Raw Score		Average Percent Correct
	MC	OE		Mean	Standard Deviation	
LAL	**20**	**5**	**48**	**24.53**	**6.53**	**51.10%**
Writing		1	12	5.90	1.97	49.17%
Reading	20	4	36	18.62	5.22	51.72%
Working with Text	11	1	15	7.79	2.74	51.93%
Analyzing Text	9	3	21	10.83	2.93	51.57%
Math	**30**	**3**	**39**	**22.17**	**7.45**	**56.85%**
Number and Numerical Operation	9	0	9	5.55	2.01	61.67%
Geometry and Measurement	7	1	10	5.85	2.01	58.50%
Patterns and Algebra	7	1	10	4.98	2.39	49.80%
Data Analysis, Probability, and Discrete Mathematics	7	1	10	5.80	2.52	58.00%
Problem Solving	12	3	21	10.20	4.81	48.57%

Table 11.3.3
Grade 7 - Means and Standard Deviations for Number Correct Raw Score for the 2007 NJ ASK Operational Forms

	Number of Items		Number of Possible Points	Raw Score		Average Percent Correct
	MC	OE		Mean	Standard Deviation	
LAL	**20**	**5**	**48**	**25.05**	**6.79**	**52.19%**
Writing		1	12	5.90	2.03	49.17%
Reading	20	4	36	19.15	5.41	53.19%
Working with Text	13	0	13	8.82	2.75	67.85%
Analyzing Text	7	4	23	10.33	3.14	44.91%
Math	**30**	**3**	**39**	**17.80**	**7.66**	**45.64%**
Number and Numerical Operation	7	1	10	4.77	2.62	47.70%
Geometry and Measurement	9	0	9	4.07	2.05	45.22%
Patterns and Algebra	7	1	10	4.66	2.23	46.60%
Data Analysis, Probability, and Discrete Mathematics	7	1	10	4.30	2.08	43.00%
Problem Solving	27	3	36	15.66	7.15	43.50%

11.4 Scale Score Distributions by Content Area and Grade

Descriptive statistics for scale scores and percentage distributions of students' performance levels are summarized in Table 11.4.1 by content area and grade. For LAL, students that were flagged as "LAL void" or "LAL not present" (LAL Scale Score marked as missing) were removed. For mathematics, students that were flagged as "Math void" or "math not present" (Math Scale Score marked as missing) were removed. For all test forms, scale scores have a range of 100 to 300. A student is classified as Partially Proficient (PP) if his/her scale score is lower than 200. A student is classified as Advanced Proficient (AP) if his/her scale score is 250 or higher. The other students are classified as Proficient (P).

11.5 Scale Score Distributions by Demographic Group

Descriptive statistics of scale scores and percentage distributions of students' performance levels by demographic groups are summarized in Tables 11.5.1 through 11.5.6 by content area and grade. For LAL, students that were flagged as "LAL void" or "LAL not present" (LAL Scale Score marked as missing) were removed. For mathematics, students that were flagged as "Math void" or "math not present" (Math Scale Score marked as missing) were removed. Students flagged as "Special Education" (SpeEdu) or "Limited English Proficient" (LEP) were removed from the "General Education" (GenEdu) group.

Table 11.4.1
Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by Content Area and Grade

Test	Grade	Form	N	Mean	StdDev	Min	Max	%PP	%P	%AP
LAL	5	Operational	101265	223.07	21.46	112	300	12.88	75.90	11.23
	6	Operational	103146	212.09	29.05	100	300	12.72	63.09	24.20
	7	Operational	104350	217.22	25.20	100	300	12.30	67.86	19.83
	6	Alternative	118	204.09	22.64	127	265	2.54	65.25	32.20
	7	Alternative	8	169.00	27.96	108	206	12.50	87.50	0.00
	5	Braille	6	209.00	21.00	179	234	66.67	33.33	0.00
	6	Braille	9	221.78	30.33	169	265	22.22	66.67	11.11
	7	Braille	3	184.00	43.49	150	233	33.33	66.67	0.00
	5	Large Print	99	206.95	22.38	158	264	5.05	61.62	33.33
	6	Large Print	103	189.77	29.88	100	265	4.85	42.72	52.43
	7	Large Print	96	199.66	28.56	110	272	3.13	53.13	43.75
Math	5	Operational	101941	228.57	32.83	100	300	29.45	54.75	15.80
	6	Operational	103689	220.11	29.98	100	300	18.99	60.04	20.98
	7	Operational	104902	211.05	34.08	100	300	15.84	50.48	33.68
	6	Alternative	116	218.63	23.64	156	269	10.34	68.97	20.69
	7	Alternative	8	158.50	14.41	137	183	100.00	0.00	0.00
	5	Braille	6	207.67	29.93	154	244	83.33	16.67	0.00
	6	Braille	9	216.22	26.79	174	250	11.11	55.56	33.33
	7	Braille	3	179.67	44.46	154	231	33.33	66.67	0.00
	5	Large Print	97	209.40	36.70	113	288	15.46	51.55	32.99
	6	Large Print	100	197.48	31.64	100	296	6.00	47.00	47.00
	7	Large Print	92	197.41	33.72	134	300	11.96	31.52	56.52

Table 11.5.1
LAL Grade 5 - Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by Demographic Groups for the 2007 NJ ASK Operational Forms

DemoGrp	N	Mean	STD	Min	Max	%PP	%P	%AP
GenEdu	81702	227.36	19.59	134	300	5.32	79.29	15.40
SpeEdu	15945	205.16	19.51	112	300	35.93	61.79	2.28
CurrLEP	4070	204.03	19.82	112	295	38.30	59.31	2.38
Female	49574	225.88	21.93	112	300	9.27	74.31	16.42
Male	51591	220.40	20.63	112	300	13.08	77.44	9.48
Asian	8144	234.13	21.12	160	300	3.82	69.92	26.26
Black	17240	210.86	19.20	112	300	24.21	72.30	3.49
Hispanic	18167	213.02	19.75	112	300	20.99	74.49	4.52
Native	78	219.41	20.71	179	295	10.26	83.33	6.41
Pacific	182	226.28	21.04	160	295	6.59	77.47	15.93
White	56903	228.42	19.73	147	300	5.22	78.29	16.49
Other	551	220.02	21.78	160	295	14.88	74.05	11.07
NonEcnDis	72100	227.95	20.29	112	300	5.96	77.38	16.66
EcnDis	29165	210.99	19.41	112	300	24.25	72.23	3.52
NonMigrant	101233	223.08	21.46	112	300	11.22	75.90	12.88
Migrant	32	204.41	15.94	165	243	28.13	71.88	0.00
NonTitle1	81355	226.56	20.75	134	300	7.68	76.87	15.45
Title 1	19910	208.79	18.14	112	300	25.74	71.90	2.36
Sect 504A	15249	204.93	18.91	112	300	35.45	62.72	1.83
Sect 504B	15074	204.90	18.94	112	300	35.57	62.60	1.82
Sect 504C	547	205.02	19.46	134	269	35.47	62.71	1.83
Sect 504D	14460	204.52	18.77	112	300	36.15	62.15	1.70

Table 11.5.2
LAL Grade 6 - Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by Demographic Groups for the 2007 NJ ASK Operational Forms

DemoGrp	N	Mean	STD	Min	Max	%PP	%P	%AP
GenEdu	83729	218.54	25.02	100	300	15.39	69.22	15.38
SpeEdu	16171	184.65	28.62	100	300	61.77	37.02	1.21
CurrLEP	3569	181.18	30.17	100	294	65.87	32.92	1.20
Female	50116	215.87	28.84	100	300	20.84	63.23	15.94
Male	52915	208.53	28.80	100	300	27.36	62.96	9.68
Asian	7828	227.65	27.85	100	300	10.59	61.01	28.40
Black	17830	194.61	28.39	100	300	47.17	49.75	3.08
Hispanic	18551	198.29	28.19	100	300	40.98	54.94	4.08
Native	136	208.38	29.79	109	283	27.21	60.29	12.50
Pacific	174	221.85	30.01	100	294	17.82	60.92	21.26
White	57926	219.79	25.13	100	300	13.54	70.14	16.31
Other	701	209.86	31.32	100	300	28.67	59.06	12.27
NonEcnDis	73382	218.86	26.58	100	300	15.51	67.91	16.57
EcnDis	29764	195.40	28.14	100	300	45.60	51.18	3.22
NonMigrant	103119	212.10	29.05	100	300	24.19	63.09	12.72
Migrant	27	190.37	33.56	132	251	48.15	48.15	3.70
NonTitle1	84744	216.46	27.61	100	300	18.23	66.72	15.06
Title 1	18402	191.95	26.97	100	300	51.70	46.37	1.93
Sect 504A	15392	185.48	28.23	100	300	60.41	38.47	1.11
Sect 504B	15329	185.51	28.31	100	300	60.34	38.50	1.17
Sect 504C	572	189.17	29.48	100	274	54.02	43.88	2.10
Sect 504D	14391	184.63	28.16	100	300	61.72	37.30	0.98

Table 11.5.3
LAL Grade 7 - Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by Demographic Groups for the 2007 NJ ASK Operational Forms

DemoGrp	N	Mean	STD	Min	Max	%PP	%P	%AP
GenEdu	85020	222.67	22.03	100	300	11.66	73.49	14.85
SpeEdu	16456	193.96	23.65	100	300	54.95	43.90	1.15
CurrLEP	3099	188.35	26.41	100	280	61.89	37.37	0.74
Female	50747	220.25	24.66	100	300	16.30	68.85	14.85
Male	53496	214.37	25.36	100	300	23.15	66.94	9.91
Asian	7579	231.13	24.52	100	300	7.77	64.88	27.35
Black	18512	202.48	23.76	100	300	39.97	57.05	2.98
Hispanic	18398	205.37	24.35	100	300	33.79	62.24	3.97
Native	105	215.35	22.86	155	280	20.95	66.67	12.38
Pacific	181	223.14	24.96	122	300	12.71	71.27	16.02
White	59020	223.80	22.22	100	300	10.64	73.45	15.91
Other	555	209.94	25.83	131	280	29.73	60.72	9.55
NonEcnDis	75190	222.86	23.47	100	300	12.49	71.54	15.97
EcnDis	29160	202.67	23.64	100	300	38.77	58.38	2.85
NonMigrant	104314	217.22	25.19	100	300	19.82	67.87	12.31
Migrant	36	191.33	31.48	122	252	52.78	41.67	5.56
NonTitle1	84821	221.18	24.11	100	300	14.54	70.73	14.72
Title 1	19529	200.01	22.45	100	297	42.81	55.39	1.80
Sect 504A	15523	194.90	23.71	100	300	53.31	45.46	1.23
Sect 504B	15450	194.95	23.74	100	300	53.21	45.53	1.26
Sect 504C	503	197.00	26.37	100	272	48.71	48.31	2.98
Sect 504D	14296	193.93	23.50	100	297	54.85	44.15	0.99

Table 11.5.4
Mathematics Grade 5 - Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by Demographic Groups for the 2007 NJ ASK Operational Forms

DemoGrp	N	Mean	STD	Min	Max	%PP	%P	%AP
GenEdu	81693	234.47	29.64	100	300	9.87	56.07	34.06
SpeEdu	15961	204.20	34.04	100	300	40.30	49.33	10.36
CurrLEP	4756	205.61	34.77	100	300	39.07	49.29	11.65
Female	49858	228.66	31.57	100	300	14.99	56.18	28.83
Male	51978	228.54	33.96	100	300	16.53	53.39	30.07
Asian	8273	249.48	31.10	108	300	5.28	38.63	56.09
Black	17272	210.39	31.63	100	300	31.78	55.92	12.30
Hispanic	18586	215.67	31.82	100	300	26.09	57.92	15.99
Native	78	224.59	31.56	119	300	14.10	61.54	24.36
Pacific	184	234.93	30.43	136	298	10.33	54.89	34.78
White	56986	235.29	29.55	100	300	9.09	55.71	35.20
Other	562	222.16	35.87	108	300	22.06	52.85	25.09
NonEcnDis	72397	235.12	30.83	100	300	10.12	54.14	35.73
EcnDis	29544	212.51	32.03	100	300	29.72	56.23	14.04
NonMigrant	101903	228.58	32.82	100	300	15.80	54.75	29.46
Migrant	38	202.47	30.77	128	279	34.21	60.53	5.26
NonTitle1	81789	233.25	31.56	100	300	11.73	54.30	33.97
Title 1	20152	209.56	30.96	100	300	32.36	56.57	11.07
Sect 504A	15260	204.04	33.38	100	300	40.02	50.18	9.80
Sect 504B	15085	203.97	33.46	100	300	40.18	49.98	9.84
Sect 504C	545	204.43	34.66	100	300	40.37	49.54	10.09
Sect 504D	14475	203.42	33.23	100	300	40.61	50.11	9.28

Table 11.5.5
Mathematics Grade 6 - Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by Demographic Groups for the 2007 NJ ASK Operational Forms

DemoGrp	N	Mean	STD	Min	Max	%PP	%P	%AP
GenEdu	83715	225.66	27.96	100	300	13.82	63.62	22.56
SpeEdu	16156	195.64	26.33	100	300	53.06	43.41	3.53
CurrLEP	4150	201.07	27.89	100	300	43.23	51.20	5.57
Female	50364	220.24	29.17	100	300	19.95	61.64	18.41
Male	53204	220.01	30.71	100	300	21.91	58.55	19.54
Asian	7931	242.03	29.68	123	300	6.34	48.33	45.33
Black	17811	201.35	26.33	100	300	42.64	52.32	5.04
Hispanic	18902	208.97	26.84	100	300	31.13	60.90	7.97
Native	137	213.45	27.11	150	269	27.01	64.23	8.76
Pacific	177	231.01	27.45	162	296	10.17	61.58	28.25
White	58003	226.54	27.68	100	300	12.96	63.76	23.27
Other	728	215.84	31.25	123	300	26.79	56.04	17.17
NonEcnDis	73642	226.21	28.96	100	300	14.35	61.58	24.07
EcnDis	30047	205.16	27.05	100	300	37.22	56.25	6.54
NonMigrant	103654	220.12	29.98	100	300	20.97	60.04	18.99
Migrant	35	199.91	30.75	134	254	45.71	45.71	8.57
NonTitle1	85115	223.99	29.44	100	300	16.66	61.21	22.13
Title 1	18574	202.35	25.71	109	300	40.75	54.66	4.59
Sect 504A	15377	196.41	26.26	100	300	51.42	44.98	3.60
Sect 504B	15312	196.41	26.29	100	300	51.51	44.87	3.62
Sect 504C	569	200.75	28.62	109	300	44.64	49.56	5.80
Sect 504D	14380	195.81	26.12	100	300	52.26	44.35	3.39

Table 11.5.6
Mathematics Grade 7 - Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by Demographic Groups for the 2007 NJ ASK Operational Forms

DemoGrp	N	Mean	STD	Min	Max	%PP	%P	%AP
GenEdu	84965	217.29	32.38	100	300	25.29	55.78	18.93
SpeEdu	16429	183.15	27.20	100	300	71.36	26.33	2.31
CurrLEP	3740	189.82	29.32	100	297	61.60	34.12	4.28
Female	51026	210.33	32.69	100	300	33.42	52.28	14.31
Male	53761	211.77	35.32	100	300	33.86	48.82	17.32
Asian	7695	236.63	35.09	111	300	11.93	47.88	40.19
Black	18485	190.02	28.36	100	300	59.63	37.02	3.34
Hispanic	18800	197.06	29.13	100	300	49.44	45.13	5.43
Native	107	207.54	32.20	146	300	40.19	46.73	13.08
Pacific	184	222.59	30.02	154	297	16.30	61.96	21.74
White	59065	218.80	31.90	100	300	23.30	56.77	19.93
Other	566	202.23	33.18	111	300	45.05	43.82	11.13
NonEcnDis	75450	217.91	33.45	100	300	25.43	54.26	20.31
EcnDis	29452	193.45	28.96	100	300	54.80	40.80	4.40
NonMigrant	104864	211.06	34.08	100	300	33.66	50.49	15.85
Migrant	38	186.39	26.19	137	240	65.79	34.21	0.00
NonTitle1	85168	215.77	33.68	100	300	27.86	53.38	18.76
Title 1	19734	190.65	27.65	100	300	58.78	37.99	3.23
Sect 504A	15499	184.32	27.62	100	300	69.67	27.74	2.59
Sect 504B	15422	184.35	27.60	100	300	69.60	27.80	2.61
Sect 504C	502	184.78	30.67	111	289	68.53	28.29	3.19
Sect 504D	14276	183.30	27.03	100	300	71.04	26.72	2.25

11.6 Scale Score Distributions by District Factor Groups (DFG)

New Jersey has a long history of using District Factor Groups (DFG)[25] in the analysis of assessment results. The statistical method for developing DFGs includes U.S. Census data and has been improved over the years to provide a better measure of social economic status (SES). The DFGs included in the tables below are A, B, CD, DE, FG, GH, I, J, R, and V. Note that group "R" in the District Factor Groups is charter schools and group "V" is vocational schools.

Descriptive statistics for scale scores and percentage distributions of students' performance levels by DFG are summarized in Tables 11.6.1 and 11.6.2 by content area and grade. For LAL, students that were flagged as "LAL void" or "LAL not present" (LAL Scale Score marked as missing) were removed. For mathematics, students that were flagged as "Math void" or "math not present" (Math Scale Score marked as missing) were removed.

[25] For more information on DFGs, see the following link: http://www.state.nj.us/njded/finance/sf/dfg.shtml.

Table 11.6.1
LAL - Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by DFG

Grade	DFG	N	Mean	StdDev	Min	Max	%PP	%P	%AP
5	A	16105	208.73	19.69	112	300	28.44	68.54	3.01
	B	9937	216.03	19.77	112	300	17.25	76.90	5.85
	CD	9538	219.49	19.70	134	300	12.39	79.45	8.16
	DE	13002	223.57	19.70	147	300	8.21	79.95	11.84
	FG	12556	225.41	19.05	155	300	6.47	81.20	12.34
	GH	14051	228.79	20.31	147	300	5.49	76.95	17.56
	I	19689	232.44	19.94	147	300	3.43	74.27	22.30
	J	4359	236.28	19.41	169	300	1.77	71.30	26.93
	R	1621	211.34	18.70	147	280	24.12	72.30	3.58
6	A	16198	191.43	29.65	100	300	51.48	45.75	2.77
	B	10366	201.30	27.00	100	300	36.61	58.55	4.84
	CD	9917	208.55	26.36	100	300	26.70	65.17	8.13
	DE	13415	213.40	25.28	100	300	19.66	69.82	10.53
	FG	13349	215.16	25.07	100	300	18.28	69.71	12.01
	GH	14106	220.36	26.68	100	300	14.88	66.87	18.25
	I	19419	225.40	25.38	100	300	9.42	67.65	22.93
	J	4244	230.56	24.32	118	300	5.75	65.50	28.75
	R	1697	198.20	28.04	109	300	43.61	51.74	4.66
	V	13	223.69	20.64	194	258	7.69	69.23	23.08
7	A	16767	199.20	24.50	100	300	44.77	52.88	2.35
	B	10591	208.05	23.38	100	300	29.67	65.40	4.94
	CD	9813	213.69	23.00	100	300	21.60	70.89	7.51
	DE	13582	217.74	22.37	100	300	16.92	72.79	10.29
	FG	13765	220.83	22.12	100	300	13.28	74.33	12.39
	GH	13705	224.48	23.06	100	300	10.57	72.13	17.30
	I	19787	228.95	22.15	100	300	7.21	70.47	22.33
	J	4281	233.27	21.96	122	300	4.72	67.27	28.01
	R	1602	207.73	22.30	131	297	32.58	62.23	5.18
	V	21	223.57	15.76	191	250	4.76	90.48	4.76

Table 11.6.2
Mathematics - Descriptive Statistics for Scale Scores and Percentage Distributions of Students' Performance Levels by DFG

Grade	DFG	N	Mean	StdDev	Min	Max	%PP	%P	%AP
5	A	16364	210.47	33.73	100	300	33.12	52.86	14.02
	B	10075	218.70	31.81	100	300	23.14	58.00	18.86
	CD	9578	223.30	30.67	100	300	18.22	59.54	22.24
	DE	13061	229.67	30.09	108	300	12.91	58.29	28.80
	FG	12587	230.85	29.55	100	300	11.93	58.65	29.42
	GH	14110	236.32	30.36	100	300	8.92	54.43	36.65
	I	19743	240.91	29.87	100	300	6.76	50.30	42.94
	J	4379	246.76	29.66	108	300	4.73	43.85	51.43
	R	1629	212.24	34.32	108	300	33.52	51.26	15.22
6	A	16396	202.92	27.68	100	300	40.77	53.21	6.01
	B	10456	210.20	27.16	100	300	30.09	60.75	9.16
	CD	9957	216.22	27.92	100	300	23.04	63.31	13.65
	DE	13434	220.61	27.93	100	300	17.96	64.58	17.46
	FG	13390	222.61	27.44	100	300	16.00	65.06	18.94
	GH	14176	226.51	28.87	123	300	14.38	61.10	24.52
	I	19468	232.44	28.66	109	300	10.10	58.56	31.34
	J	4269	239.66	27.91	109	300	6.09	53.13	40.78
	R	1699	205.27	28.42	109	300	39.38	52.09	8.53
	V	13	235.77	29.13	181	284	7.69	53.85	38.46
7	A	16973	190.85	29.21	100	300	58.81	37.21	3.98
	B	10700	199.85	30.74	100	300	45.96	46.29	7.75
	CD	9865	205.31	30.93	100	300	38.63	51.26	10.11
	DE	13611	209.96	31.24	100	300	33.26	53.72	13.02
	FG	13814	213.89	31.18	100	300	27.53	57.12	15.35
	GH	13750	219.96	33.37	100	300	23.25	54.90	21.85
	I	19820	226.57	32.97	111	300	17.26	54.76	27.98
	J	4297	234.21	33.38	111	300	11.94	51.92	36.14
	R	1606	195.75	31.58	100	300	53.99	39.29	6.72
	V	21	212.81	22.44	161	262	14.29	80.95	4.76

PART 12: RELIABILITY

12.1 Introduction

The New Jersey Department of Education is required by federal law to ensure that the instruments it uses to measure student achievement for school accountability provide reliable results. This section shows that results of the 2007 NJ ASK for grades 5–7 measure student achievement in a reliable manner. The size of the measurement error associated with test scores is reasonable and can be taken into account when interpreting the scores for individual students.

12.2 Reliability and Measurement Error

Reliable student test scores, like other reliable measurements, are consistent. More specifically, measurement components are consistent with each other. Results of the components vary, but they do so within tolerable limits. In general, measurement error and reliability are inversely related. When measurement error is large, reliability is small. Increasing reliability by minimizing measurement error is an important goal in the construction of any test.

The NJ ASK assessments, like many other standardized achievement tests, were designed under the assumptions of Classical Test Theory (CTT). This approach builds on the notion of an ideal, error-free or true measurement score. Any observed measurement, such as test score X, is defined as a composite of true score T and its associated error:

$$X = T + \text{error}$$

Estimating the size of the measurement error associated with the true score is the key to estimating reliability. Errors in measurement can result from any of a multitude of factors, including environmental factors (e.g., testing conditions) and examinee factors (e.g., fatigue, stress). Feldt and Brennan (1989)[26] note, "Quantification of the consistency and inconsistency in examinee performance constitutes the essence of reliability analysis" (p. 105). CTT provides a means for this quantification of examinee inconsistency (i.e., measurement error).

The definitions or assumptions in CTT lead to several important properties. For example, it can be demonstrated that

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2,$$

or observed score variance equals the sum of true score variance plus error variance. The relationships among variance terms (i.e., $\sigma_x^2, \sigma_t^2, \sigma_e^2$) are critical to a more thorough understanding of important CTT concepts, including reliability and the standard error of

[26] Feldt and Brennan (1989). Reliability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed.). Washington, DC: American Council on Education.

measurement. For example, CTT reliability is defined as the correlation between observed scores on parallel forms, which is equal to

$$\rho_{x_1 x_2} = \sigma_t^2 \Big/ \sigma_x^2 .$$

Reliability in CTT is thus conceptualized as true score variance divided by observed score variance. With just a few algebraic steps, the CTT definition of the standard error of measurement (SEM) can be shown as

$$\sigma_e = \sigma_x \sqrt{1 - \rho_{x_1 x_2}} .$$

Although the conceptualization of reliability and SEM is relatively straightforward, issues underlying the estimation of reliability are not. Reliability can be estimated via the correlation of scores on parallel forms or from test-retest data, or it can be estimated from a single test administration using any one of a variety of techniques (e.g., Brown, 1910; Cronbach, 1951; Kuder & Richardson, 1937)[27]. A very popular technique for estimating reliability from a single test administration is Cronbach's coefficient alpha.

12.3 Test Metrics and Units of Analysis

The NJ ASK quantifies student achievement on three different metrics: number correct raw score, IRT scale, and performance score. While it is the knowledge and skills of individual students that are measured, student scores are aggregated and disaggregated into various units (e.g., school by grade, student group by grade, school, district, and state). Measurement error specific to each metric and each unit of analysis is taken into account when results are reported and accountability decisions are made. It is the responsibility of test developers to maximize reliability and minimize error by (1) identifying likely sources of error; (2) controlling the conditions of error; (3) estimating the size of error and/or level of reliability; and (4) reporting the estimates by metric and unit of analysis.

12.4 Sources of Measurement Error

The scoring of student responses to multiple-choice items is done electronically. Scoring error may result from improper coding and extraneous marks on scannable response sheets. The size of this sort of error is usually small and is controlled though proper test administration procedures, including instructions on how to fill out response sheets and how to erase extraneous markings. The test contractor, who manages the scoring, also uses procedures to minimize this error.

[27] Brown, W. (1910). Some experimental results in the correlation of mental abilities. *British Journal of Psychology, 3*, 296–322. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika, 2*, 151–160.

Measurement Incorporated (MI) performs a *cusp check*, or multiple choice verification process, for any student whose bubbling errors, if corrected, would give them a passing score. This involves identifying the affected multiple-choice (MC) answer pages and physically reviewing each one for each student in this group. Two of the most common types of errors are use of pen rather than pencil and double-bubbling, which often turns out to be just a bad erasure. If the student's responses are clear to the human eye, as opposed to the machine's eye, the correct response is recorded and the score is changed.

Open-ended items are susceptible to scoring error due to ambiguity in scoring rubrics as well as to differences among raters. Rubrics must be written to balance generality and specificity, covering the range of student responses, while at the same time allowing raters to easily identify the response characteristics distinguishing each score category. To minimize error due to raters, the test contractor thoroughly trains raters and monitors the scoring process. Only raters who meet the contractor's criteria for consistent scoring during training are retained as scorers. The contractor monitors scoring by routinely computing and recording inter-rater agreement.

12.5 Evidence of Raw Score Internal Consistency

Consistency of individual student performance was estimated using coefficient alpha. As previously noted, coefficient alpha is conceptualized as the proportion of total raw score variance that may be attributed to a student's true score variance. Ideally, more score variance should be attributable to true test scores than to measurement error. Alpha is an appropriate index of internal consistency for use on untimed tests such as NJ ASK.

Separate analyses were performed for each grade level and content area. Both multiple-choice and open-ended items scores were used in the computations. Coefficient alpha can be interpreted as a lower bound to reliability and was estimated using the following formula:

$$\alpha_{\text{Cronbach}} = \frac{n}{n-1}[1 - \frac{\sum_{i=1}^{n} \sigma_{Y_i}^2}{\sigma_X^2}],$$

where n is the number of items, $\sigma_{Y_i}^2$ is the variance of item i, and σ_X^2 is the variance of total score. SEM can be interpreted as "the square root of the average of the person-specific error variances of all examinees who participated in the reliability estimation experiment" (Traub, 1994, p. 114[28]). SEMs were calculated using the following formula:

$$SEM = S_X \sqrt{1 - \alpha_{\text{Cronbach}}},$$

where S_X is the standard deviation of observed total scores. Table 12.5.1 summarizes coefficient alpha and SEMs by content and form. Tables 12.5.2 through 12.5.5 summarize coefficient alpha and SEMs of content clusters by test.

[28] Traub, R. E. (1994). Reliability for the social sciences, v3. Thousand Oaks, Sage.

Table 12.5.1
Summary of Coefficient Alpha and SEM by Grade and Content Area for the 2007 NJ ASK Operational Forms

Test	Form*	Grade	Ncount	Cronbach Alpha	SEM
LAL	OP	5	101265	0.87	2.03
	OP	6	103146	0.82	2.74
	OP	7	104350	0.83	2.82
Math	OP	5	101941	0.84	2.65
	OP	6	103689	0.87	2.73
	OP	7	104902	0.87	2.74

*OP: Operational Test; N-counts were insufficient to produce values for Braille and Large Print.

Table 12.5.2
Grade 5 - Coefficient Alpha and Standard Error Measurement for Clusters for the 2007 NJ ASK Operational Forms

	Number of Items		Number of Possible Points	Alpha	SEM
	MC	OE			
LAL	**20**	**5**	**41**	**0.87**	**2.03**
Writing		1	5		
Reading	20	4	36	0.86	1.92
Working with Text	12	1	16	0.75	1.42
Analyzing Text	8	3	20	0.76	1.33
Math	**30**	**3**	**39**	**0.84**	**2.65**
Number and Numerical Operation	7	1	10	0.63	1.42
Geometry and Measurement	9	0	9	0.49	1.12
Patterns and Algebra	7	1	10	0.49	1.40
Data Analysis, Prob., & Discrete Math.	7	1	10	0.62	1.42
Problem Solving	16	3	25	0.74	2.28

Table 12.5.3
Grade 6 - Coefficient Alpha and Standard Error Measurement for Clusters for the 2007 NJ ASK Operational Forms

	Number of Items		Number of Possible points	Alpha	SEM
	MC	OE			
LAL	**20**	**5**	**48**	**0.82**	**2.74**
Writing		1	12		
Reading	20	4	36	0.84	2.07
Working with Text	11	1	15	0.69	1.53
Analyzing Text	9	3	21	0.77	1.41
Math	**30**	**3**	**39**	**0.87**	**2.73**
Number and Numerical Operation	9	0	9	0.64	1.21
Geometry and Measurement	7	1	10	0.55	1.35
Patterns and Algebra	7	1	10	0.61	1.48
Data Analysis, Prob., & Discrete Math.	7	1	10	0.67	1.45
Problem Solving	12	3	21	0.79	2.20

Table 12.5.4
Grade 7 - Coefficient Alpha and Standard Error Measurement for Clusters for the 2007 NJ ASK Operational Forms

	Number of Items		Number of Possible points	Alpha	SEM
	MC	OE			
LAL	**20**	**5**	**48**	**0.83**	**2.82**
Writing		1	12		
Reading	20	4	36	0.84	2.14
Working with Text	13	0	13	0.74	1.41
Analyzing Text	7	4	23	0.73	1.62
Math	**30**	**3**	**39**	**0.87**	**2.74**
Number and Numerical Operation	7	1	10	0.63	1.59
Geometry and Measurement	9	0	9	0.61	1.29
Patterns and Algebra	7	1	10	0.63	1.35
Data Analysis, Prob., & Discrete Math.	7	1	10	0.63	1.27
Problem Solving	27	3	36	0.86	2.66

12.6 Evidence Supporting Rater Reliability

Tables 12.6.1, 12.6.2, and 12.6.3 show the percentages of writing tasks and open-ended items scored with exact agreement, adjacent agreement, and resolution needed by grade level and content area. The score rubrics used for raters had a score range of 0 to 5 for the grade 5 writing prompt, and 0 to 6 for the grade 6 and 7 writing prompt. For grade 6 and 7, the writing scores were doubled in data analyses and score reporting. The rubrics had score points that ranged from 0 to 4 for the LAL open-ended items and from 0 to 3 for the mathematics open-ended items. There were no half points assigned for the writing and open-ended items.

Ten percent (10%) of the writing and open-ended responses were read by a second rater. The purpose of the second-reading for the writing and open-ended responses was to investigate the consistency between raters for the 2007 NJ ASK. For grade 5 LAL, over 79% of the responses were assigned a score by a second rater that was in exact agreement with the first rater. Another 20% of the second ratings were assigned an adjacent score by a second rater. An adjacent score is a score assigned by the second rater that is no more than ±1 score point from the score assigned by the first rater. For grade 5 mathematics, over 91% of the responses were assigned a score by a second rater that was in exact agreement with the first rater. The results for grades 6 and 7 were comparable, with the exception of grade 7 LAL where the percentage of exact agreement was lower than the other three grades.

Table 12.6.1

Grade 5 - Consistency between Rater Scoring for the Writing Task and Open-Ended Items

	% Raters in Exact Agreement	% Raters in Adjacent Agreement	% Resolution Needed
LAL All	79.4	20.3	0.2
Writing	73.9	25.4	0.6
Reading All	80.8	19.0	0.2
Open-ended 1	80.8	19.0	0.2
Open-ended 2	82.2	17.8	0.0
Open-ended 3	79.8	20.0	0.2
Open-ended 4	80.5	19.2	0.2
Math All	91.5	8.2	0.2
Open-ended 1	94.0	5.8	0.0
Open-ended 2	90.3	9.4	0.2
Open-ended 3	90.3	9.4	0.4

Table 12.6.2

Grade 6 - Consistency between Rater Scoring for the Writing Task and Open-Ended Items

	% Raters in Exact Agreement	% Raters in Adjacent Agreement	% Resolution Needed
LAL All	76.5	23.2	0.3
Writing	64.8	33.8	1.2
Reading All	79.5	20.5	0.1
Open-ended 1	84.0	16.0	0.0
Open-ended 2	84.0	15.8	0.2
Open-ended 3	75.4	24.6	0.0
Open-ended 4	74.4	25.6	0.2
Math All	88.7	10.3	0.8
Open-ended 1	93.3	6.0	0.6
Open-ended 2	90.6	9.0	0.4
Open-ended 3	82.2	16.0	1.4

Table 12.6.3
Grade 7 - Consistency between Rater Scoring for the Writing Task and Open-Ended Items

	% Raters in Exact Agreement	% Raters in Adjacent Agreement	% Resolution Needed
LAL All	67.2	30.9	1.6
Writing	63.1	34.6	2.0
Reading All	68.2	30.0	1.5
Open-ended 1	74.1	25.2	0.6
Open-ended 2	71.3	27.6	0.6
Open-ended 3	65.4	33.0	1.4
Open-ended 4	62.0	34.2	3.4
Math All	92.3	3.5	0.2
Open-ended 1	95.7	4.2	0.2
Open-ended 2	96.1	3.8	0.2
Open-ended 3	97.0	2.6	0.2

12.7 Conditional Estimate of Error at Each Cut-Score

The 2007 NJ ASK grades 5, 6, and 7 raw score cut scores and the corresponding conditional standard error of measurement (CSEM) are summarized in Table 12.7.1. The CSEM is calculated as:

$$CSEM = (SD_{obs} / SD_{adj'd}) * SE_{theta}$$

where: SD_{obs} is the observed standard deviation for number correct raw score,
$SD_{adj'd}$ is an estimate of the "true" sample standard deviation, and
SE_{theta} is the standard error for theta.

Table 12.7.1
Raw Score Cut Scores with Conditional Standard Error of Measurement by Content Area and Grade Level for the 2007 NJ ASK Operational Forms

		Language Arts Literacy		Mathematics	
		Proficient	Advanced Proficient	Proficient	Advanced Proficient
Grade 5	Cut score	17	30	18	29
	CSEM	2.51	1.45	2.71	2.61
Grade 6	Cut score	21	32	16	30
	CSEM	2.56	2.09	2.74	2.58
Grade 7	Cut score	20	33	14	27
	CSEM	2.56	2.08	2.72	2.69

12.8 Reliability of Classifications

Reliability indices for proficiency classifications (kappa) were computed with the BB-CLASS program (Brennan, 2004)[29] that is based on the beta-binomial model. Coefficient kappa is given by:

$$\kappa = \frac{\varphi - \varphi_c}{1 - \varphi_c},$$

where φ is the probability of a consistent classification and φ_c is the probability of a consistent classification by chance. A classification consistency index can be regarded as the percentage of examinees that would be assigned to, hypothetically, the same achievement level if the same test was administered a second time or an equivalent test was administered under the same conditions. Coefficient kappa depends on test score variability, test length, and achievement levels as well as the position of cut-scores (Huynh, 1976)[30].

Coefficients kappa and consistency classification indices are summarized in Table 12.8.1. Kappa and the consistency indexes were estimated for three achievement levels (PP, P, and AP) and two achievement levels (Pass or Fail). A student is regarded as "Pass" if one's achievement level is "P" or "AP" and as "Fail" if one's achievement level is "PP". The latter classification accuracy is directly related to determining the accuracy of Proficiency classifications for NCLB. These classification accuracy values were generally in the mid- to upper-eighties.

Table 12.8.1
Consistency Indexes for Performance Levels
for the 2007 NJ ASK Operational Forms

Test	Grade	Three Achievement Levels (PP, P, and AP)			Two Achievement Levels (Pass [P+AP], Fail [PP])		
		Cut-scores	Kappa	φ	Cut-score	Kappa	φ
LAL	5	17,30	0.34	75%	17	0.71	94%
	6	21,32	0.43	69%	21	0.64	87%
	7	20,33	0.43	73%	20	0.67	90%
Math	5	18,29	0.52	71%	18	0.64	90%
	6	16,30	0.57	76%	16	0.64	88%
	7	14,27	0.58	75%	14	0.64	85%

[29] Brennan, R. L. (2004). Manual for BB-CLASS: A computer program that uses the beta-binomial model for classification consistency and accuracy (version 1). CASMA Research Report 9. Iowa City, IA.
[30] Huynh, H. (1976). On the reliability of decisions in domain-referenced testing, *Journal of Educational Measurement*. 13:253-264.

PART 13: VALIDITY

13.1 Introduction

The *Standards for Educational and Psychological Testing* states, "Ultimately, the validity of an intended interpretation of test scores relies on all the available evidence relevant to the technical quality of a testing program. This includes evidence of careful test construction; adequate score reliability; appropriate test administration and scoring; accurate score scaling, equating, and standard setting; and careful attention to fairness for all examinees," (page 17).[31] While this section summarizes evidence supporting claims as to the validity of NJ ASK performance scores, many parts of this Technical Report provide appropriate evidence for validity. Some of this evidence is cross-referenced below for added convenience. Given the procedural and empirical evidence available and the rationale presented below, valid performance standards-based interpretations and uses of the scores are generally supported.

The following begins with a review of important federal statutes requiring the NJ ASK for grades 5–7 and goes on to explain the purposes and intended uses of performance test scores, suggesting the value implications of performance scores for schools, teachers, students, and parents. Content-related evidence supporting validity is presented in terms of the adequacy and appropriateness of the state content standards and the representation of the content standards on the tests. Then, validity evidence based on the internal structure of NJ ASK is provided through a correlational analysis of NJ ASK content clusters with each other. Reference to specific Standards within the *Standards for Educational and Psychological Testing* are provided where appropriate.

13.2 Federal Authority for School Accountability

NJ ASK performance standards scores for grades 5-7 reflect student achievement with respect to performance standards Partially Proficient, Proficient, and Advanced Proficient. The United States Department of Education bases accountability on school achievement of Adequate Yearly Progress (AYP) in reading/language arts and mathematics. AYP is set in terms of percentages of all students, and all student groups, scoring Proficient or above on the NJ ASK mathematics and LAL assessments.

13.3 Purpose and Intended Uses of Test Performance Scores[32]

The NJ ASK was developed for the following purposes and uses:

[31] American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for Educational and Psychological Testing*. Washington: APA.

[32] Standard 1.2 – The test developer should set forth clearly how test scores are intended to be interpreted and used. The population(s) for which a test is appropriate should be clearly delimited, and the construct that the test is intended to assess should be clearly described (page 17).

- NJ ASK tests scores provide an indication of student progress toward achieving the knowledge and skills identified in the New Jersey's Core Curriculum Content Standards (CCCS) and the tests fulfill the requirements under the No Child Left Behind (NCLB) Act;
- Annual school-improvement planning for instructional programs and other efforts directed toward the increased achievement of students in grade-, school-, or district-level groups; school improvement to reduce performance-gaps between disaggregated minority/majority groups such as racial, gender, disability, poverty, and limited English proficiency;
- Student, teacher, and parent information concerning the academic performance levels of individual students. Such information should be considered with other information, including other assessments and teachers judgment.

The valid interpretation and appropriate use of NJ ASK scores is supported in a variety of ways, including the training and consultation provided by personnel of the New Jersey Department of Education and publications such as the *New Jersey Assessment of Skills & Knowledge Spring 2007 Test Coordinator Manual Grades 5–7* and this Technical Report. The training and documentation provided to test users helps them better administer, understand, and use test score results.

13.4 NJ ASK Test Scores

The NJ ASK for grades 5–7 are scaled in several ways: raw score points, Item Response Theory (IRT), and performance standard level (based on scale-score cuts). New Jersey actively promotes the use of performance level results, reporting them annually on each content test at the student, school, district and state levels. Individual student and average scale scores are also used, but should play a secondary role, generally interpreted with reference to their distance from performance-score cut points. Test results are reported for students as a whole as well as by student group including gender, ethnicity, disability, use of accommodations, English language proficiency, migrant status, and poverty. Scores are reported to schools and districts in the annually published reports (see Part 8: Reporting).

NJ ASK performance scores indicate that an individual student performs at the Partially Proficient, Proficient, and Advanced Proficient level in a content area. Performance standard descriptions associated with each level provide details of the performance that students have met or exceeded. No stakes for students or teachers are attached by the state to student-level scores. Teachers are counseled to interpret individual student scores only in the context of other assessment results and their own experience.

13.5 Content-Related Evidence of Validity[33]

Baker and Linn (2002)[34] suggest that "Two questions are central in the evaluation of content aspects of validity. Is the definition of the content domain to be assessed adequate and appropriate? Does the test provide an adequate representation of the content domain the test is intended to measure?" (p. 6). The following two sections help answer these two very important questions and also address Standard 1.6 of the *Standards for Educational and Psychological Testing*.

13.6 Appropriateness of Content Definition

In 1996, the New Jersey State Board of Education adopted the New Jersey Core Curriculum Content Standards, an ambitious framework for educational reform in the State's public schools. New Jersey's standards were created to improve student achievement by clearly defining what all students should know and be able to do at the end of thirteen years of public education. Since the adoption of those standards, the NJDOE has continuously engaged in discussion with educators, business representatives, and national experts about the impact of the standards on classroom practices. To assist teachers and curriculum specialists in aligning curriculum with the standards, the NJDOE provided local school districts with a curriculum framework for each content area. The frameworks provided classroom teachers and curriculum specialists with sample teaching strategies, adaptations, and background information relevant to each of the content areas.

The State Board wisely required that the standards be reviewed and revised every five years. The review process, begun in May 2001, involved teachers, school administrators, students, parents, and representatives from business, higher education, and the community. In addition, several content areas were reviewed by Achieve, Inc., and the Council of Chief State School Officers (CCSSO). In response to this unprecedented review, the 2004 New Jersey Core Curriculum Content Standards provide the level of specificity and depth of content that will better prepare students for post secondary education and employment. The standards are based on the latest research in each of the content areas and identify the essential core of learning for all students.

Since the adoption of the original 1996 New Jersey Core Curriculum Content Standards (CCCS), the New Jersey State Board of Education approved administrative code that implements all aspects of standards-based reform. N.J.A.C. 6A:8 requires districts to align all curriculum to the standards; ensure that teachers provide instruction according to the standards; ensure student performance is assessed in each content area; and provide teachers with opportunities for professional development that focuses on the standards.

---

[33] Standard 1.6 – When the validation rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified in reference to the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified (page 18).

[34] Baker, E. L., & Linn, R. L. (2002). Validity Issues for Accountability Systems. Center for the Study of Evaluation. Technical Report 585, Los Angeles, CA.

13.7 Adequacy of Content Representation

Adequacy of the content representation of the NJ ASK is critically important because the tests must provide an indication of student progress toward achieving the knowledge and skills identified in the CCCS, and the tests must fulfill the requirements under NCLB.

Adequate representation of the content domains defined in the CCCS is assured through use of a test blueprint and a responsible test construction process. New Jersey performance standards, as well as the CCCS, are taken into consideration in the writing of multiple-choice and open-ended items and open-ended rubric development. Each test must align with and proportionally represent the sub domains of the test blueprint. Evidence to support the above was given in Part 2, Test Development Process, and Part 4, Item Analysis. Tables 2.3.1 through 2.3.12 in Part 2 provide a comparison of target test construction maps to actual test maps for LAL and mathematics. Inspection of these tables confirms that the target number of items for each sub domain was achieved. As a representative example of this match, Table 2.3.1 is repeated below, with the addition of the actual counts as found in Table 2.3.2 (in parentheses).

Table 2.3.1
Test Construction Map for Grade 5 Language Arts Literacy NJ ASK

Text types/Strand	Reading Selections	MC (Number of Items)	OE (Number of Items)	WT (Number of Items)	Total Points
Picture Prompt		0 (0)	0 (0)	1 (1)	5 (5)
Narrative	1 (1)				
AT		4-6 (6)	0-2 (1)	0 (0)	8-12 (10)
WT		4-6 (4)	0-2 (1)	0 (0)	8-12 (8)
Everyday Text	1 (1)				
AT		2-6 (2)	0-2 (2)	0 (0)	8-12 (10)
WT		4-8 (8)	0-2 (0)	0 (0)	8-12 (8)
Total Items		**20 (20)**	**4 (4)**	**1 (1)**	
Total Points		**20 (20)**	**16 (16)**	**5 (5)**	**41 (41)**

The contractor strives to equitably represent the CCCS on each test by balancing sub-domain coverage on each test, by proportionally representing items corresponding to Partially Proficient, Proficient, and Advanced Proficient performance categories on each test, and by matching item format to the requirements of the content and standards descriptions.

13.8 Validity Evidence Based on the Internal Structure of NJ ASK[35]

Because the NJ ASK testing program assesses student performance in several content areas using a variety of testing methods, it is important to study the pattern of relationships among the content areas and testing methods. One method for studying patterns of relationships to provide evidence supporting the inferences made from test scores is the multi-trait, multi-method matrix. Tables 13.8.1, 13.8.2, and 13.8.3 summarize Pearson correlation coefficients among test content domains and clusters by grade level. The correlations between clusters within a content area were generally found to be higher than the correlations between clusters across the two content areas.

Table 13.8.1
Grade 5 - Correlation Coefficients among Content Domains and Clusters

	Writing	Reading	LAL1	LAL 2	Math	Math 1	Math 2	Math 3	Math 4	Math 5
Writing	1.00									
Reading	0.50	1.00								
Interpreting (LAL1)	0.45	0.94	1.00							
Analyzing (LAL2)	0.49	0.93	0.75	1.00						
Math	0.43	0.69	0.65	0.64	1.00					
Number (Math 1)	0.38	0.60	0.57	0.56	0.85	1.00				
Geometry (Math 2)	0.30	0.48	0.45	0.44	0.70	0.49	1.00			
Patterns (Math 3)	0.32	0.51	0.48	0.47	0.78	0.55	0.42	1.00		
Data Analysis (Math 4)	0.37	0.61	0.57	0.56	0.85	0.63	0.48	0.53	1.00	
Prob. Solving (Math 5)	0.41	0.66	0.62	0.61	0.96	0.83	0.64	0.78	0.79	1.00

[35] Standard 1.11 – If the rationale for a test use or interpretation depends on premises about the relationships among parts of the test, evidence concerning the internal structure of the test should be provided.

Standard 1.12 – When interpretation of subscores, score differences, or profiles is suggested, the rationale and relative evidence in support of such interpretation should be provided. Where composite scores are developed, the basis and rationale for arriving at the composites should be given.

Table 13.8.2
Grade 6 - Correlation Coefficients among Content Domains and Clusters

	Writing	Reading	LAL1	LAL 2	Math	Math 1	Math 2	Math 3	Math 4	Math 5
Writing	1.00									
Reading	0.56	1.00								
Interpreting (LAL1)	0.48	0.91	1.00							
Analyzing (LAL2)	0.55	0.93	0.69	1.00						
Math	0.49	0.69	0.64	0.63	1.00					
Number (Math 1)	0.41	0.57	0.53	0.52	0.83	1.00				
Geometry (Math 2)	0.38	0.54	0.51	0.49	0.81	0.58	1.00			
Patterns (Math 3)	0.42	0.61	0.57	0.55	0.85	0.63	0.60	1.00		
Data Analysis (Math 4)	0.43	0.57	0.52	0.52	0.84	0.59	0.55	0.59	1.00	
Prob. Solving (Math 5)	0.47	0.66	0.62	0.59	0.96	0.79	0.78	0.80	0.81	1.00

Table 13.8.3
Grade 7 - Correlation Coefficients among Content Domains and Clusters

	Writing	Reading	LAL1	LAL 2	Math	Math 1	Math 2	Math 3	Math 4	Math 5
Writing	1.00									
Reading	0.58	1.00								
Interpreting (LAL1)	0.49	0.91	1.00							
Analyzing (LAL2)	0.57	0.93	0.69	1.00						
Math	0.51	0.66	0.62	0.60	1.00					
Number (Math 1)	0.43	0.56	0.53	0.51	0.88	1.00				
Geometry (Math 2)	0.40	0.51	0.48	0.46	0.83	0.64	1.00			
Patterns (Math 3)	0.46	0.61	0.57	0.55	0.85	0.66	0.62	1.00		
Data Analysis (Math 4)	0.44	0.56	0.52	0.52	0.84	0.65	0.59	0.63	1.00	
Prob. Solving (Math 5)	0.50	0.65	0.61	0.59	1.00	0.88	0.82	0.85	0.84	1.00

13.9 Additional Evidence for Validity of NJ ASK

Validity evidence related to other Standards is listed below:

Standard 1.5[36]

- The composition of the sample of examinees from which validity evidence was obtained is described in detail in Part 11, including major relevant sociodemographic characteristics. This information is imbedded within the Tables of Part 11. These tables also provide descriptive statistics for number correct raw score and for scale scores. Statistics include N-counts, means, standard deviations, minimum and maximum values, and a variety of data disaggregations, including student demographic group and District Factor Group (DFG).

Standard 1.7[37]

- Standard setting procedures, including the selection process and the characteristics of judges, is described in detail in Part 3.
- The 2007 NJ ASK open-ended questions and writing responses at grades 5, 6, and 7 required hand scoring by Measurement Incorporated (MI) personnel. The processes of selecting and training scorers, reading and scoring papers, and monitoring scoring are described in detail in Part 6.

Standard 1.13[38]

- The conditions under which the data were collected are described in Part 5. Information about the administration of NJ ASK is available in the *New Jersey Assessment of Skills & Knowledge Spring 2007 Test Coordinator Manual Grades 5–7*.

[36] Standard 1.5 - The composition of any sample of examinees from which validity evidence is obtained should be described in as much detail as is practical, including major relevant sociodemographic and developmental characteristics.

[37] Standard 1.7 – When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications, and experience, of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.

[38] Standard 1.13 - When validity evidence includes statistical analyses of test results, either alone or together with data on other variables, the conditions under which the data were collected should be described in enough detail that users can judge the relevance of the statistical findings to local conditions. Attention should be drawn to any features of a validation data collection that are likely to differ from typical operational testing conditions and that could plausibly influence test performance.

APPENDIX 2-1
CHECKLIST FOR FORMS DEVELOPMENT

Table 2-1
Checklist for Forms Development

Item Data
Target Sum Rasch is between 495 and 500. If Target is not met, manager is made aware.
As many items as possible have a p-value above 0.35 and below 0.90
As many items as possible have a pt. bis above 0.20
No item was used as a sample item.

Item Pool
No passages are from the operational NJPASS
No writing prompts are from the Operational NJPASS (with the possible exception of the breach form writing prompt).
The NJPASS FT linking items are not used across grades.

Item Distribution
Item standards are distributed equally throughout the test
There are a variety of indicators assessed in each standard
MC items are generally in passage order, and OE items are at the end of the passage sets. WT items are in the appropriate places.
Answer key distribution is nearly equal between answer choices: A B C D
There are NOT more than 2 MC items in a row with the same answer.

Name, Gender, and Ethnicity Distributions
Check gender distribution (number of passages or prompts which have a male and/or female): Male Female Both
Check ethnicity distribution (number of passages or prompts): Caucasian_Hispanic Asian_African American Other_____
There are NOT two or more items in the same session that have similar contexts.
There are NOT two or more items with similar answers or answer choices.
Sample items and test items do NOT clue each other.
Items do NOT have any fairness or sensitivity related to the names and contexts of the items.

APPENDIX 3-1
DEMOGRAPHIC BACKGROUND OF PANELISTS PARTICIPATING IN STANDARD SETTING

Table 3-1
Demographic Background of PLD Panelists by Content Area/Grade Level

		LAL 5	LAL6	LAL 7	Math 5	Math 6	Math 7
Gender	M	2	0	1	2	2	1
	F	7	9	9	7	8	9
Ethnicity	W	9	8	8	9	7	7
	AA		1	2		2	2
Years Exp.	1-4	0	2	1	0	1	0
	5-10	4	1	2	3	1	1
	11-14	1	1	3	3	1	2
	15-19	2	0	0	2	4	1
	20-24	2	1	0	1	3	3
	25-30	0	4	1	0	0	1
	30+	0	0	3	1	0	1
DFG	A	1	3	1	1	2	4
	B	1	0	1	2	1	0
	CD	1	1	0	1	1	2
	DE	1	1	3	1	2	0
	FG	1	0	2	1	1	2
	GH	2	2	2	1	2	1
	I	2	2	0	3	1	1
	other	0	0	1	0	0	0
Region	N	3	4	6	6	3	6
	S	6	5	4	4	7	4
	Urban	2	5	1	3	2	5
Position	Teacher	5	6	6	7	7	5
	Adm.	4	3	4	3	3	5
Degree	B.A.	1	3	3	6	2	5
	M.A.	8	6	7	4	8	3
	Ph.D.	0	0	0	0	0	1

Table 3-2
Demographic Background of Phase 1 Panelists by Content Area and Grade Level

		Language Arts Literacy			Mathematics			Overall	
Grade		5	6	7	5	6	7	N	%
Gender	M	3	3	1	4	4	4	19	16
	F	16	16	19	16	16	16	99	84
Ethnicity	W	16	18	17	14	13	12	90	76
	AA	2	1	2	2	6	4	17	14
	H/L	1	0	1	2	0	3	7	6
Years Exp.	1-4	2	0	2	1	2	2	9	8
	5-10	6	4	6	6	3	6	31	26
	11-14	4	4	4	5	4	6	27	23
	15-19	1	4	1	0	3	2	11	9
	20-24	2	2	4	1	2	0	11	9
	25-30	3	3	3	2	4	2	17	14
	30+	1	2	0	0	2	2	7	6
DFG	A	1	1	3	3	2	3	13	11
	B	3	1	3	2	2	1	12	10
	CD	3	4	2	2	3	2	16	14
	DE	3	6	1	2	2	3	17	14
	FG	1	0	5	4	4	4	18	15
	GH	3	1	1	2	4	3	14	12
	I	4	3	2	3	2	2	16	14
	J	0	2	1	1	1	1	6	5
	CS	1	1	2	1	0	1	6	5
Region	N	9	8	13	8	10	12	60	51
	S	10	11	7	12	10	8	58	49
Position	Teacher	18	15	17	14	17	15	96	81
	Adm.	1	4	3	6	3	5	22	19
Degree	B.A.	5	4	5	10	9	6	39	33
	M.A.	14	15	15	9	11	13	77	65
	Ph.D.	0	0	0	1	0	1	2	2

APPENDIX 3-2
PERFORMANCE LEVEL DESCRIPTORS

New Jersey Assessment of Skills and Knowledge (NJ ASK)
Performance Level Descriptors
Language Arts Literacy
Grade 5

Proficient

The student performing at the proficient level demonstrates an ability to analyze and critique text, synthesizing details in order to build understanding. The student constructs meaning recognizing literary elements and figurative language. The student recognizes author's purpose and applies prior knowledge to draw conclusions.

As a proficient writer, the student develops a central focus and organizes and connects ideas with relevant details. The student exhibits some variety in word choice and sentence structure, attempts narrative techniques, and gives some evidence of transitions while incorporating basic writing mechanics.

Advanced Proficient

As a reader, the fifth grade student is able to synthesize details in order to make connections and generate new ideas. The student utilizes literary elements and author's purpose to analyze text.

As an advanced proficient writer, the student establishes and maintains a strong central focus and elaborates supporting details to convey ideas effectively. The student will include narrative techniques using fluid transitions, strong appropriate word choice and sentence variety to purposefully engage the reader.

New Jersey Assessment of Skills and Knowledge (NJ ASK)
Performance Level Descriptors
Language Arts Literacy
Grade 6

Proficient

The student performing at the proficient level demonstrates an understanding of the literal and inferential aspects of the text. Students activate prior knowledge and sustain comprehension by questioning, clarifying and predicting. Students at this level extrapolate relevant details to determine main idea, infer word meaning, in context, and identify the author's purpose. The proficient student recognizes organizational patterns and textual conventions.

As a proficient writer, the student develops a single focus and supporting details within an organizational structure. The student writes for a variety of purposes keeping audience in mind. Students at this level provide support for opinions and conclusions using textual and literary elements effectively.

Advanced Proficient

As a reader, the sixth grade advanced proficient student extends meaning by making conscious connections to text, author, self, and others. The student synthesizes information and draws insightful conclusions. Students at this level demonstrate an understanding of the central theme.

As an advanced proficient writer, in addition to consistently demonstrating the qualities outlined for a proficient student, the advanced proficient student develops a logical progression of ideas in a fluent and cohesive voice. Students at this level take appropriate compositional risks.

New Jersey Assessment of Skills and Knowledge (NJ ASK)
Performance Level Descriptors
Language Arts Literacy
Grade 7

Proficient

Seventh-grade students performing at the proficient level are able to interact with a variety of texts. Proficient students demonstrate an understanding of literal and inferential levels, while making relevant use of text to support or explain self-generated text. Students are able to identify central themes, supporting details and organizational structures of text. Students monitor their understanding of text and identify purpose for reading. Students can interpret textual conventions and extrapolate and synthesize information from text.

Seventh-grade students proficient in writing are able to develop a central theme, supporting details, and an organizational structure. Students at this level establish a purpose for writing and provide support for opinions and conclusions. Students demonstrate control of textual and literary elements.

Advanced Proficient

As an advanced proficient reader, in addition to consistently demonstrating the qualities outlined for a proficient student, the advanced student identifies abstract themes while analyzing and evaluating text. The advanced student utilizes prior knowledge to extend aspects of the text.

As an advanced proficient writer, in addition to consistently developing a central theme, supporting details and organizational structure, the advanced student demonstrates sophisticated use of textual and literary elements. The advanced student effectively uses compositional risks.

New Jersey Assessment of Skills and Knowledge (NJ ASK)
Performance Level Descriptors
Mathematics
Grade 5

Proficient

The student performing at the proficient level demonstrates the ability to recognize, understand, and apply basic mathematical concepts, skills, and terminology to theoretical and real world situations. The student will be able to perform basic computational procedures, apply geometric properties and spatial relationships, interpret data and graphs, apply the concepts and methods of discrete mathematics, and use basic algebraic concepts and processes. The student will also be able to infer, reason, and estimate while problem solving. The student will demonstrate flexibility in selecting a successful process or strategy. The student will demonstrate basic understanding of mathematical concepts through written expression.

Advanced Proficient

The student performing at the advanced proficient level will consistently demonstrate the qualities outlined for proficient performance. In addition, the student will analyze methods for appropriateness, synthesize processes, and evaluate mathematical relationships. The student will demonstrate conceptual understanding by consistently providing clear and complete explanations. The student abstracts concepts for use in other applications and successfully forms conjectures.

New Jersey Assessment of Skills and Knowledge (NJ ASK)
Performance Level Descriptors
Mathematics
Grade 6

Proficient

The student performing at the proficient level in mathematics will be able to demonstrate evidence of and communicate conceptual understanding of procedural and analytical skills. The student applies mathematical skills and knowledge to theoretical and real world situations. In addition, the student makes connections across mathematical domains. The student at this level understands and applies appropriate standard numerical operations and estimations—an understanding sufficient for problem solving in practical situations. The student will be able to determine the reasonableness of an answer. The student understands and applies basic geometric concepts including properties, measurement, and spatial relationships. The student clearly interprets data and graphs, determines probabilities, applies the concepts and methods of discrete mathematics, and uses basic algebraic concepts and processes.

Advanced Proficient

The student performing at the advanced proficient level in mathematics will consistently demonstrate the qualities for proficient performance. In addition, the student at the advanced level demonstrates the use of abstract thinking and provides explanations that are consistently clear and thorough. The student will support a logical efficient method to solving the problem. The student consistently makes accurate inferences and predictions. The student supports responses by using appropriate mathematical terminology. The student successfully analyzes and draws appropriate inferences from data.

New Jersey Assessment of Skills and Knowledge (NJ ASK)
Performance Level Descriptors
Mathematics
Grade 7

Proficient

A proficient student demonstrates a conceptual understanding of mathematical knowledge, procedures, skills, and processes across the four content standards. A student will demonstrate an ability to:

- Identify, recognize, and compare different representations of numbers and numerical operations.
- Identify, describe, classify, apply, and solve problems involving geometry, spatial sense, and measurement.
- Recognize, identify, and extend simple patterns and algebraic representations of theoretical and real-world problems.
- Model situations, solve problems, analyze, and draw appropriate inferences from data.
- Understand fundamental concepts of probability and discrete mathematics.

Advanced Proficient

An advanced proficient student consistently demonstrates the qualities outlined for proficient performance. In addition, the student consistently demonstrates the ability to:

- Think abstractly using inductive and deductive reasoning.
- Use a variety of problem solving strategies.
- Extrapolate information; form and support conclusions through clear and thorough explanation(s).
- Assess reasonableness of the solution(s).

APPENDIX 7-1
RAW SCORE TO SCALE SCORE TABLES

LAL Grade 5			
2007 Operational			
Raw Sc.	Scale Sc.	Theta	S.E.
0	112	-5.923	1.832
1	134	-4.701	1.015
2	147	-3.973	0.734
3	155	-3.526	0.616
4	160	-3.189	0.549
5	165	-2.912	0.507
6	169	-2.671	0.477
7	173	-2.453	0.456
8	176	-2.254	0.439
9	179	-2.066	0.427
10	182	-1.889	0.417
11	185	-1.718	0.409
12	187	-1.553	0.403
13	190	-1.392	0.399
14	192	-1.234	0.397
15	194	-1.077	0.396
16	197	-0.920	0.396
17	200	-0.762	0.398
18	201	-0.602	0.402
19	204	-0.439	0.407
20	206	-0.270	0.414
21	209	-0.095	0.423
22	212	0.089	0.435
23	215	0.284	0.449
24	219	0.493	0.467
25	222	0.721	0.489
26	226	0.974	0.516
27	231	1.257	0.550
28	237	1.581	0.590
29	243	1.956	0.636
30	250	2.395	0.688
31	259	2.906	0.743
32	269	3.503	0.804
33	280	4.210	0.880
34	295	5.069	0.973
35	300	6.087	1.029
36	300	7.089	0.949
37	300	7.879	0.839
38	300	8.559	0.828
39	300	9.311	0.923
40	300	10.366	1.169
41	300	11.820	1.918

LAL Grade 6			
2007 Operational			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-6.146	1.837
1	100	-4.912	1.021
2	100	-4.177	0.737
3	100	-3.727	0.616
4	109	-3.392	0.547
5	118	-3.118	0.503
6	125	-2.881	0.473
7	132	-2.668	0.451
8	138	-2.472	0.436
9	144	-2.287	0.424
10	149	-2.111	0.415
11	155	-1.942	0.408
12	160	-1.778	0.403
13	164	-1.617	0.398
14	169	-1.460	0.395
15	174	-1.306	0.392
16	178	-1.153	0.390
17	182	-1.001	0.389
18	186	-0.851	0.388
19	190	-0.700	0.388
20	194	-0.550	0.389
21	200	-0.398	0.390
22	201	-0.245	0.392
23	205	-0.090	0.396
24	208	0.068	0.400
25	212	0.230	0.405
26	216	0.396	0.411
27	220	0.568	0.418
28	224	0.746	0.427
29	228	0.933	0.437
30	233	1.129	0.449
31	238	1.337	0.463
32	250	1.559	0.479
33	251	1.797	0.497
34	258	2.054	0.517
35	265	2.332	0.538
36	274	2.634	0.561
37	283	2.962	0.586
38	294	3.321	0.613
39	300	3.717	0.646
40	300	4.160	0.687
41	300	4.666	0.737
42	300	5.255	0.799
43	300	5.951	0.871
44	300	6.774	0.939
45	300	7.691	0.971
46	300	8.664	1.010
47	300	9.834	1.199
48	300	11.327	1.930

LAL Grade 7			
2007 Operational			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-5.793	1.841
1	100	-4.551	1.026
2	110	-3.806	0.741
3	122	-3.352	0.618
4	131	-3.016	0.547
5	139	-2.742	0.502
6	145	-2.506	0.471
7	150	-2.295	0.450
8	155	-2.100	0.435
9	160	-1.915	0.424
10	165	-1.738	0.417
11	169	-1.567	0.412
12	173	-1.399	0.407
13	177	-1.235	0.404
14	181	-1.073	0.401
15	184	-0.913	0.398
16	187	-0.755	0.396
17	191	-0.599	0.394
18	194	-0.445	0.392
19	197	-0.292	0.391
20	200	-0.139	0.390
21	203	0.013	0.390
22	206	0.165	0.391
23	209	0.319	0.393
24	212	0.474	0.395
25	215	0.631	0.399
26	218	0.792	0.404
27	221	0.958	0.410
28	225	1.129	0.418
29	229	1.307	0.427
30	233	1.494	0.438
31	237	1.692	0.451
32	242	1.902	0.466
33	250	2.128	0.484
34	252	2.371	0.503
35	258	2.633	0.523
36	265	2.917	0.543
37	272	3.223	0.562
38	280	3.548	0.579
39	288	3.892	0.594
40	297	4.253	0.608
41	300	4.632	0.624
42	300	5.034	0.643
43	300	5.463	0.668
44	300	5.932	0.703
45	300	6.462	0.757
46	300	7.103	0.856
47	300	8.025	1.106
48	300	9.383	1.884

Math Grade 5			
20007 Operational			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-5.507	1.853
1	100	-4.234	1.047
2	100	-3.448	0.768
3	100	-2.955	0.647
4	108	-2.584	0.576
5	119	-2.279	0.529
6	128	-2.018	0.495
7	136	-1.785	0.470
8	143	-1.574	0.450
9	150	-1.379	0.434
10	156	-1.196	0.421
11	162	-1.024	0.410
12	168	-0.859	0.401
13	174	-0.702	0.394
14	179	-0.549	0.387
15	184	-0.402	0.381
16	189	-0.259	0.376
17	194	-0.118	0.372
18	200	0.019	0.368
19	204	0.153	0.365
20	208	0.286	0.363
21	213	0.416	0.361
22	217	0.546	0.360
23	222	0.675	0.359
24	226	0.804	0.360
25	230	0.934	0.362
26	235	1.066	0.365
27	239	1.201	0.369
28	244	1.339	0.376
29	250	1.483	0.384
30	254	1.635	0.395
31	259	1.796	0.409
32	265	1.970	0.427
33	271	2.163	0.452
34	279	2.382	0.486
35	287	2.641	0.535
36	298	2.966	0.610
37	300	3.414	0.739
38	300	4.158	1.028
39	300	5.405	1.844

Math Grade 6			
2007 Operational			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-5.469	1.860
1	100	-4.179	1.057
2	109	-3.377	0.776
3	123	-2.874	0.653
4	134	-2.498	0.579
5	142	-2.192	0.529
6	150	-1.932	0.493
7	156	-1.703	0.465
8	162	-1.496	0.444
9	167	-1.308	0.426
10	172	-1.132	0.412
11	177	-0.967	0.400
12	181	-0.811	0.391
13	185	-0.662	0.382
14	189	-0.518	0.376
15	193	-0.379	0.370
16	200	-0.244	0.366
17	201	-0.112	0.362
18	204	0.018	0.359
19	208	0.147	0.357
20	211	0.274	0.356
21	215	0.401	0.356
22	219	0.527	0.356
23	222	0.654	0.357
24	226	0.781	0.358
25	229	0.910	0.360
26	233	1.040	0.362
27	237	1.173	0.366
28	241	1.309	0.371
29	245	1.449	0.378
30	250	1.596	0.388
31	254	1.751	0.402
32	258	1.920	0.420
33	263	2.106	0.445
34	269	2.320	0.480
35	276	2.573	0.529
36	284	2.891	0.604
37	296	3.328	0.730
38	300	4.055	1.018
39	300	5.285	1.836

Math Grade 7			
2007 Operational			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-5.390	1.843
1	100	-4.141	1.031
2	111	-3.386	0.750
3	126	-2.919	0.628
4	137	-2.571	0.557
5	146	-2.287	0.510
6	154	-2.044	0.477
7	161	-1.829	0.451
8	167	-1.635	0.431
9	172	-1.456	0.416
10	178	-1.289	0.403
11	183	-1.131	0.392
12	188	-0.981	0.383
13	192	-0.838	0.375
14	200	-0.700	0.368
15	201	-0.567	0.361
16	205	-0.439	0.356
17	209	-0.314	0.351
18	213	-0.192	0.347
19	217	-0.073	0.343
20	221	0.044	0.341
21	225	0.159	0.340
22	228	0.275	0.340
23	232	0.391	0.343
24	236	0.510	0.347
25	240	0.632	0.353
26	244	0.759	0.361
27	250	0.893	0.371
28	253	1.036	0.384
29	257	1.189	0.399
30	262	1.356	0.417
31	268	1.538	0.437
32	274	1.739	0.460
33	281	1.963	0.486
34	289	2.213	0.515
35	297	2.496	0.551
36	300	2.827	0.604
37	300	3.248	0.705
38	300	3.913	0.971
39	300	5.057	1.793

LAL Grade 5			
2007 Braille			
Raw Sc.	Scale Sc.	Theta	S.E.
0	112	-5.517	1.710
1	133	-4.376	0.948
2	145	-3.696	0.686
3	152	-3.277	0.576
4	158	-2.962	0.514
5	162	-2.701	0.475
6	166	-2.474	0.447
7	170	-2.269	0.428
8	173	-2.080	0.413
9	176	-1.903	0.401
10	179	-1.735	0.393
11	182	-1.572	0.386
12	185	-1.415	0.381
13	187	-1.261	0.378
14	190	-1.108	0.376
15	193	-0.957	0.376
16	195	-0.805	0.378
17	200	-0.651	0.380
18	201	-0.494	0.385
19	203	-0.333	0.391
20	206	-0.166	0.400
21	209	0.010	0.410
22	213	0.196	0.424
23	216	0.396	0.440
24	220	0.613	0.461
25	224	0.853	0.487
26	229	1.124	0.518
27	234	1.433	0.556
28	240	1.791	0.600
29	250	2.208	0.648
30	255	2.694	0.699
31	264	3.258	0.754
32	275	3.923	0.824
33	288	4.729	0.910
34	300	5.682	0.962
35	300	6.619	0.887
36	300	7.357	0.784
37	300	7.992	0.773
38	300	8.694	0.861
39	300	9.678	1.091
40	300	11.035	1.790

Math Grade 5			
2007 Braille / Large Print			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-4.693	1.659
1	100	-3.569	0.929
2	100	-2.886	0.677
3	113	-2.462	0.569
4	124	-2.143	0.507
5	133	-1.881	0.466
6	140	-1.655	0.437
7	147	-1.454	0.415
8	154	-1.270	0.398
9	160	-1.100	0.385
10	165	-0.941	0.374
11	171	-0.789	0.365
12	176	-0.645	0.357
13	181	-0.505	0.351
14	185	-0.371	0.345
15	190	-0.240	0.340
16	194	-0.113	0.336
17	200	0.011	0.333
18	203	0.133	0.330
19	207	0.253	0.327
20	211	0.371	0.326
21	215	0.489	0.324
22	219	0.605	0.324
23	223	0.722	0.324
24	227	0.839	0.326
25	231	0.958	0.329
26	235	1.080	0.333
27	239	1.205	0.338
28	244	1.334	0.346
29	250	1.471	0.355
30	253	1.616	0.368
31	258	1.773	0.385
32	264	1.946	0.407
33	271	2.144	0.438
34	278	2.377	0.482
35	288	2.670	0.549
36	300	3.073	0.665
37	300	3.742	0.925
38	300	4.865	1.660

Math Grade 6			
2007 Braille			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-4.327	1.581
1	113	-3.254	0.886
2	131	-2.602	0.646
3	142	-2.197	0.542
4	151	-1.894	0.482
5	158	-1.646	0.442
6	164	-1.434	0.413
7	169	-1.246	0.391
8	174	-1.076	0.374
9	178	-0.919	0.360
10	182	-0.773	0.349
11	186	-0.635	0.340
12	190	-0.503	0.332
13	193	-0.377	0.326
14	197	-0.255	0.321
15	200	-0.136	0.317
16	203	-0.020	0.314
17	206	0.094	0.311
18	209	0.206	0.310
19	213	0.318	0.309
20	216	0.429	0.309
21	219	0.540	0.309
22	222	0.652	0.310
23	225	0.764	0.311
24	228	0.877	0.313
25	232	0.992	0.316
26	235	1.110	0.320
27	239	1.231	0.326
28	242	1.358	0.334
29	246	1.493	0.345
30	250	1.638	0.361
31	255	1.798	0.382
32	260	1.982	0.412
33	266	2.199	0.454
34	273	2.472	0.518
35	283	2.847	0.626
36	300	3.471	0.873
37	300	4.525	1.574

Math Grade 7			
2007 Large Print			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-4.900	1.676
1	100	-3.765	0.937
2	121	-3.078	0.682
3	134	-2.653	0.571
4	144	-2.335	0.507
5	153	-2.076	0.465
6	160	-1.855	0.435
7	166	-1.658	0.412
8	172	-1.480	0.394
9	177	-1.315	0.381
10	182	-1.160	0.370
11	187	-1.013	0.361
12	191	-0.872	0.354
13	195	-0.736	0.349
14	200	-0.604	0.345
15	204	-0.475	0.342
16	208	-0.347	0.340
17	212	-0.220	0.339
18	216	-0.094	0.339
19	220	0.032	0.340
20	224	0.160	0.341
21	228	0.289	0.344
22	233	0.421	0.348
23	237	0.556	0.354
24	241	0.696	0.361
25	250	0.842	0.369
26	251	0.997	0.380
27	256	1.161	0.393
28	261	1.338	0.409
29	267	1.530	0.428
30	274	1.741	0.449
31	281	1.976	0.474
32	289	2.239	0.506
33	299	2.544	0.553
34	300	2.930	0.642
35	300	3.536	0.882
36	300	4.573	1.628

Math Grade 7			
2007 Braille			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-4.872	1.677
1	100	-3.735	0.939
2	122	-3.045	0.684
3	135	-2.617	0.574
4	146	-2.297	0.510
5	154	-2.035	0.467
6	161	-1.811	0.437
7	167	-1.612	0.414
8	173	-1.432	0.397
9	179	-1.265	0.383
10	184	-1.109	0.371
11	188	-0.962	0.361
12	193	-0.821	0.353
13	200	-0.687	0.346
14	201	-0.558	0.339
15	205	-0.433	0.334
16	209	-0.313	0.329
17	213	-0.196	0.324
18	216	-0.081	0.321
19	220	0.031	0.318
20	224	0.142	0.317
21	227	0.253	0.318
22	231	0.364	0.320
23	234	0.478	0.324
24	238	0.596	0.331
25	242	0.719	0.340
26	250	0.851	0.351
27	251	0.991	0.365
28	255	1.145	0.382
29	261	1.313	0.401
30	266	1.499	0.423
31	273	1.707	0.447
32	280	1.941	0.475
33	288	2.205	0.507
34	298	2.513	0.554
35	300	2.900	0.643
36	300	3.505	0.880
37	300	4.538	1.625

LAL Grade 5a			
Special Equating, removed 11, 12			
Raw Sc.	Scale Sc.	Theta	S.E.
0	127	-5.052	1.603
1	147	-3.983	0.888
2	158	-3.344	0.645
3	165	-2.948	0.543
4	170	-2.647	0.487
5	174	-2.396	0.452
6	177	-2.176	0.427
7	181	-1.977	0.409
8	184	-1.792	0.396
9	186	-1.618	0.386
10	189	-1.451	0.378
11	191	-1.290	0.373
12	194	-1.132	0.370
13	196	-0.976	0.369
14	200	-0.821	0.369
15	201	-0.664	0.372
16	203	-0.504	0.377
17	205	-0.339	0.384
18	208	-0.166	0.394
19	211	0.017	0.407
20	214	0.213	0.424
21	217	0.429	0.447
22	221	0.672	0.476
23	226	0.951	0.514
24	232	1.283	0.565
25	239	1.690	0.632
26	250	2.211	0.723
27	259	2.917	0.857
28	276	3.943	1.030
29	297	5.215	1.046
30	300	6.331	0.910
31	300	7.149	0.807
32	300	7.965	0.938
33	300	9.086	1.617

LAL Grade 7a			
Special Equating, removed 23			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-5.293	1.719
1	102	-4.133	0.957
2	120	-3.441	0.691
3	131	-3.019	0.576
4	139	-2.705	0.510
5	146	-2.450	0.469
6	152	-2.228	0.442
7	157	-2.028	0.423
8	162	-1.842	0.411
9	166	-1.666	0.402
10	171	-1.496	0.395
11	175	-1.330	0.391
12	178	-1.168	0.387
13	182	-1.009	0.384
14	185	-0.852	0.381
15	189	-0.698	0.379
16	192	-0.545	0.377
17	195	-0.394	0.375
18	200	-0.244	0.374
19	201	-0.094	0.374
20	204	0.056	0.374
21	206	0.206	0.375
22	209	0.358	0.377
23	212	0.511	0.381
24	215	0.669	0.386
25	219	0.830	0.392
26	222	0.998	0.399
27	226	1.172	0.409
28	230	1.356	0.420
29	234	1.552	0.435
30	239	1.762	0.451
31	244	1.990	0.471
32	250	2.239	0.494
33	255	2.515	0.520
34	262	2.820	0.547
35	270	3.157	0.575
36	279	3.527	0.600
37	289	3.928	0.623
38	300	4.360	0.646
39	300	4.823	0.670
40	300	5.323	0.698
41	300	5.875	0.740
42	300	6.520	0.822
43	300	7.415	1.046
44	300	8.702	1.765

LAL Grade 7b			
Special Equating, removed 24			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-5.398	1.719
1	100	-4.237	0.958
2	117	-3.542	0.693
3	129	-3.117	0.578
4	137	-2.802	0.512
5	144	-2.545	0.470
6	150	-2.324	0.441
7	155	-2.125	0.422
8	159	-1.940	0.409
9	164	-1.765	0.399
10	168	-1.597	0.393
11	172	-1.434	0.389
12	176	-1.274	0.385
13	180	-1.116	0.383
14	183	-0.960	0.381
15	186	-0.805	0.379
16	190	-0.652	0.377
17	193	-0.500	0.376
18	196	-0.349	0.375
19	200	-0.198	0.375
20	202	-0.047	0.375
21	204	0.104	0.377
22	207	0.257	0.379
23	210	0.412	0.382
24	213	0.570	0.387
25	217	0.733	0.393
26	220	0.901	0.400
27	224	1.076	0.410
28	228	1.261	0.422
29	232	1.458	0.436
30	237	1.670	0.454
31	242	1.900	0.474
32	250	2.154	0.499
33	254	2.435	0.526
34	261	2.747	0.554
35	269	3.093	0.582
36	278	3.472	0.607
37	288	3.881	0.629
38	299	4.319	0.651
39	300	4.789	0.674
40	300	5.295	0.702
41	300	5.853	0.744
42	300	6.504	0.825
43	300	7.404	1.048
44	300	8.693	1.766

LAL Grade 7c			
Special Equating, removed 23, 24			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-5.282	1.719
1	102	-4.122	0.958
2	120	-3.429	0.691
3	132	-3.005	0.576
4	140	-2.691	0.512
5	147	-2.434	0.471
6	152	-2.211	0.444
7	158	-2.009	0.426
8	163	-1.820	0.414
9	167	-1.640	0.406
10	171	-1.467	0.400
11	175	-1.297	0.396
12	179	-1.130	0.393
13	183	-0.966	0.391
14	186	-0.803	0.389
15	190	-0.642	0.387
16	193	-0.482	0.386
17	196	-0.322	0.385
18	200	-0.164	0.385
19	202	-0.004	0.386
20	205	0.156	0.388
21	209	0.319	0.391
22	212	0.484	0.396
23	215	0.655	0.402
24	219	0.831	0.410
25	223	1.015	0.420
26	227	1.210	0.434
27	231	1.419	0.450
28	236	1.646	0.471
29	241	1.896	0.497
30	250	2.177	0.529
31	255	2.498	0.567
32	263	2.870	0.612
33	274	3.303	0.660
34	286	3.803	0.705
35	300	4.365	0.743
36	300	4.981	0.772
37	300	5.640	0.799
38	300	6.368	0.860
39	300	7.319	1.066
40	300	8.632	1.774

LAL Grade 7d			
Special Equating, removed 25 (writing)			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-5.247	1.731
1	104	-4.053	0.982
2	124	-3.307	0.727
3	136	-2.830	0.618
4	146	-2.463	0.557
5	154	-2.156	0.516
6	161	-1.888	0.487
7	167	-1.646	0.465
8	172	-1.422	0.448
9	177	-1.214	0.435
10	182	-1.016	0.425
11	186	-0.827	0.416
12	190	-0.645	0.409
13	193	-0.467	0.404
14	197	-0.294	0.401
15	200	-0.122	0.399
16	203	0.048	0.399
17	207	0.219	0.400
18	210	0.391	0.402
19	213	0.566	0.407
20	217	0.746	0.414
21	221	0.933	0.423
22	225	1.130	0.434
23	229	1.338	0.448
24	234	1.562	0.466
25	239	1.805	0.488
26	250	2.074	0.514
27	252	2.373	0.544
28	260	2.709	0.577
29	269	3.086	0.611
30	279	3.511	0.648
31	291	3.989	0.689
32	300	4.533	0.738
33	300	5.169	0.805
34	300	5.944	0.903
35	300	6.995	1.113
36	300	8.384	1.800

Math Grade 5a			
Special Equating, removed 12			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-4.955	1.668
1	100	-3.809	0.943
2	100	-3.101	0.692
3	106	-2.657	0.583
4	117	-2.322	0.520
5	127	-2.047	0.477
6	135	-1.810	0.447
7	142	-1.600	0.424
8	149	-1.408	0.407
9	155	-1.231	0.393
10	161	-1.065	0.381
11	166	-0.908	0.372
12	172	-0.757	0.365
13	177	-0.612	0.359
14	182	-0.471	0.354
15	186	-0.334	0.350
16	191	-0.199	0.347
17	196	-0.066	0.344
18	200	0.065	0.343
19	205	0.195	0.342
20	210	0.326	0.343
21	214	0.457	0.344
22	219	0.589	0.346
23	223	0.723	0.349
24	228	0.860	0.354
25	233	1.002	0.359
26	238	1.148	0.366
27	243	1.300	0.375
28	250	1.460	0.386
29	254	1.631	0.400
30	260	1.817	0.419
31	267	2.025	0.447
32	275	2.266	0.488
33	285	2.565	0.554
34	298	2.973	0.668
35	300	3.648	0.928
36	300	4.775	1.662

Math Grade 5b			
Special Equating, removed 23,32			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-4.952	1.669
1	100	-3.805	0.943
2	100	-3.096	0.692
3	106	-2.650	0.584
4	118	-2.314	0.521
5	127	-2.038	0.479
6	135	-1.799	0.449
7	143	-1.587	0.426
8	149	-1.393	0.409
9	156	-1.214	0.396
10	162	-1.044	0.385
11	167	-0.884	0.376
12	173	-0.729	0.370
13	178	-0.580	0.364
14	183	-0.434	0.360
15	188	-0.291	0.357
16	193	-0.150	0.355
17	200	-0.011	0.354
18	203	0.128	0.353
19	208	0.266	0.353
20	212	0.405	0.354
21	217	0.545	0.356
22	222	0.686	0.359
23	227	0.831	0.364
24	232	0.981	0.372
25	237	1.139	0.382
26	243	1.308	0.398
27	250	1.493	0.420
28	256	1.703	0.451
29	264	1.950	0.495
30	274	2.257	0.561
31	288	2.673	0.673
32	300	3.352	0.928
33	300	4.477	1.660

Math Grade 6a			
Special Equating, removed 23,31			
Raw Sc.	Scale Sc.	Theta	S.E.
0	**SS**	-4.670	1.596
1	100	-3.561	0.908
2	104	-2.869	0.668
3	123	-2.433	0.563
4	136	-2.106	0.500
5	145	-1.840	0.458
6	152	-1.612	0.427
7	159	-1.411	0.404
8	164	-1.229	0.386
9	169	-1.062	0.372
10	174	-0.906	0.360
11	178	-0.758	0.351
12	182	-0.618	0.344
13	186	-0.482	0.338
14	190	-0.351	0.333
15	194	-0.223	0.330
16	200	-0.097	0.328
17	201	0.028	0.327
18	205	0.153	0.328
19	208	0.279	0.330
20	212	0.407	0.333
21	215	0.538	0.338
22	219	0.673	0.344
23	223	0.814	0.352
24	227	0.963	0.363
25	231	1.122	0.376
26	236	1.294	0.392
27	240	1.483	0.414
28	250	1.696	0.442
29	252	1.944	0.482
30	259	2.248	0.542
31	267	2.652	0.645
32	278	3.304	0.886
33	296	4.376	1.581

Math Grade 7a			
Special Equating, remove 23,32			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-4.874	1.677
1	100	-3.736	0.939
2	122	-3.047	0.684
3	135	-2.619	0.574
4	146	-2.299	0.510
5	154	-2.037	0.468
6	161	-1.812	0.438
7	167	-1.613	0.415
8	173	-1.431	0.398
9	179	-1.263	0.384
10	184	-1.105	0.374
11	189	-0.955	0.365
12	193	-0.811	0.359
13	200	-0.672	0.354
14	202	-0.536	0.350
15	206	-0.402	0.347
16	211	-0.270	0.346
17	215	-0.139	0.345
18	219	-0.008	0.346
19	224	0.124	0.347
20	228	0.258	0.350
21	232	0.394	0.354
22	237	0.534	0.360
23	241	0.680	0.368
24	250	0.832	0.378
25	251	0.994	0.391
26	257	1.170	0.408
27	263	1.363	0.431
28	269	1.581	0.461
29	277	1.836	0.504
30	287	2.151	0.569
31	300	2.573	0.680
32	300	3.258	0.937
33	300	4.393	1.675

Math Grade 7b			
Special Equating, removed 10			
Raw Sc.	Scale Sc.	Theta	S.E.
0	100	-4.888	1.676
1	100	-3.752	0.938
2	121	-3.063	0.683
3	135	-2.637	0.573
4	145	-2.318	0.509
5	153	-2.058	0.467
6	160	-1.834	0.436
7	167	-1.636	0.414
8	172	-1.456	0.396
9	178	-1.289	0.382
10	183	-1.133	0.371
11	188	-0.985	0.362
12	192	-0.844	0.355
13	200	-0.708	0.348
14	201	-0.577	0.343
15	205	-0.450	0.338
16	209	-0.326	0.333
17	213	-0.206	0.330
18	217	-0.087	0.327
19	221	0.030	0.326
20	224	0.147	0.325
21	228	0.263	0.327
22	232	0.382	0.331
23	236	0.505	0.337
24	240	0.633	0.346
25	244	0.769	0.358
26	250	0.916	0.373
27	254	1.075	0.390
28	259	1.251	0.410
29	265	1.446	0.432
30	272	1.664	0.457
31	279	1.906	0.483
32	288	2.179	0.514
33	297	2.494	0.559
34	300	2.886	0.646
35	300	3.493	0.881
36	300	4.528	1.626

APPENDIX 11-4
SCALE SCORE CUMULATIVE FREQUENCY DISTRIBUTIONS

LAL Grade 5

		All Students		Male	Female	White	Afr.-A.	Hisp.
Raw	Scale	Cumulative*		Cumul.	Cumul.	Cumul.	Cumul.	Cumul.
Score	Score	#	%	%	%	%	%	%
0	112	4	0.0	0.0	0.0	0.0	0.0	0.0
1	134	12	0.0	0.0	0.0	0.0	0.0	0.0
2	147	50	0.1	0.1	0.0	0.0	0.1	0.1
3	155	107	0.1	0.1	0.1	0.0	0.3	0.3
4	160	227	0.2	0.3	0.1	0.1	0.5	0.5
5	165	415	0.4	0.6	0.3	0.1	1.0	1.0
6	169	749	0.7	1.0	0.5	0.2	1.6	1.7
7	173	1161	1.2	1.5	0.8	0.4	2.6	2.6
8	176	1692	1.7	2.2	1.2	0.6	3.7	3.7
9	179	2363	2.3	3.0	1.6	0.8	5.2	5.0
10	182	3157	3.1	4.0	2.2	1.2	7.0	6.6
11	185	4039	4.0	5.0	3.0	1.5	9.0	8.2
12	187	5094	5.0	6.2	3.8	2.0	11.2	10.3
13	190	6320	6.2	7.6	4.8	2.5	13.9	12.6
14	192	7761	7.7	9.2	6.0	3.2	17.0	15.0
15	194	9488	9.4	11.1	7.6	4.2	20.5	17.9
16	197	11369	11.2	13.1	9.3	5.2	24.2	21.0
17	200	13704	13.5	15.6	11.3	6.5	28.6	24.9
18	201	16444	16.2	18.6	13.8	8.2	33.7	29.3
19	204	19438	19.2	21.6	16.6	10.3	38.8	33.8
20	206	23090	22.8	25.4	20.0	13.1	44.4	38.9
21	209	27435	27.1	30.1	23.9	16.6	50.4	44.6
22	212	32677	32.3	35.6	28.7	21.1	56.9	51.3
23	215	38727	38.2	42.1	34.1	26.7	63.6	58.5
24	219	45636	45.1	49.4	40.5	33.7	70.3	65.6
25	222	53650	53.0	57.7	48.0	42.3	76.9	72.8
26	226	62522	61.7	66.7	56.5	52.4	83.2	80.1
27	231	71747	70.9	75.7	65.8	63.3	88.9	86.2
28	237	80718	79.7	84.1	75.1	74.2	93.4	91.7
29	243	88225	87.1	90.5	83.6	83.5	96.5	95.5
30	250	94014	92.8	95.2	90.4	90.9	98.3	97.8
31	259	97776	96.6	97.9	95.2	95.6	99.4	99.1
32	269	99905	98.7	99.3	98.0	98.3	99.8	99.7
33	280	100836	99.6	99.8	99.4	99.5	99.9	99.9
34	295	101180	99.9	99.9	99.9	99.9	100.0	100.0
35	300	101265	100.0	100.0	100.0	100.0	100.0	100.0

*All cumulative distributions include students scored on the full set of items.

LAL Grade 6

		All Students		Male	Female	White	Afr.-A.	Hisp.
Raw	Scale	Cumulative*		Cumul.	Cumul.	Cumul.	Cumul.	Cumul.
Score	Score	#	%	%	%	%	%	%
0	100	152	0.2	0.2	0.1	0.0	0.4	0.3
4	109	324	0.3	0.4	0.2	0.1	0.7	0.7
5	118	571	0.6	0.8	0.3	0.2	1.3	1.3
6	125	901	0.9	1.2	0.6	0.3	2.1	1.9
7	132	1394	1.4	1.8	0.9	0.4	3.2	2.9
8	138	1976	1.9	2.5	1.3	0.6	4.6	4.0
9	144	2646	2.6	3.3	1.8	0.8	6.2	5.3
10	149	3450	3.3	4.2	2.4	1.2	8.0	6.8
11	155	4461	4.3	5.4	3.2	1.6	10.3	8.5
12	160	5629	5.5	6.7	4.1	2.0	13.1	10.5
13	164	6933	6.7	8.2	5.2	2.6	15.9	12.9
14	169	8583	8.3	10.0	6.6	3.3	19.5	15.8
15	174	10436	10.1	12.0	8.1	4.2	23.1	18.9
16	178	12543	12.2	14.3	9.9	5.3	27.2	22.4
17	182	15075	14.6	17.0	12.0	6.8	31.9	26.5
18	186	17853	17.3	20.0	14.5	8.5	36.6	30.9
19	190	21206	20.6	23.5	17.5	10.8	41.9	35.7
20	194	24958	24.2	27.4	20.8	13.5	47.2	41.0
21	200	29405	28.5	32.1	24.7	17.2	52.9	46.8
22	201	34325	33.3	37.2	29.2	21.4	58.8	52.9
23	205	39782	38.6	42.8	34.1	26.4	64.6	59.2
24	208	45988	44.6	49.2	39.7	32.6	70.7	65.4
25	212	52538	50.9	55.8	45.8	39.4	76.0	71.5
26	216	59306	57.5	62.4	52.3	46.9	81.0	77.1
27	220	66210	64.2	69.1	59.0	54.8	85.5	82.3
28	224	72963	70.7	75.4	65.8	62.7	89.3	86.9
29	228	79318	76.9	81.2	72.3	70.4	92.4	90.7
30	233	85074	82.5	86.2	78.6	77.5	95.0	93.7
31	238	90028	87.3	90.3	84.1	83.7	96.9	95.9
32	250	93992	91.1	93.6	88.6	88.7	98.0	97.5
33	251	97168	94.2	96.0	92.3	92.6	98.9	98.7
34	258	99499	96.5	97.7	95.2	95.6	99.4	99.3
35	265	100990	97.9	98.7	97.1	97.4	99.7	99.7
36	274	101938	98.8	99.3	98.3	98.6	99.9	99.8
37	283	102517	99.4	99.7	99.1	99.3	99.9	99.9
38	294	102844	99.7	99.9	99.6	99.7	100.0	100.0
39	300	103146	100.0	100.0	100.0	100.0	100.0	100.0

*All cumulative distributions include students scored on the full set of items.

LAL Grade 7

		All Students		Male	Female	White	Afr.-A.	Hisp.
Raw	Scale	Cumulative*		Cumul.	Cumul.	Cumul.	Cumul.	Cumul.
Score	Score	#	%	%	%	%	%	%
0	100	38	0.0	0.1	0.0	0.0	0.1	0.1
2	110	95	0.1	0.1	0.1	0.0	0.2	0.2
3	122	185	0.2	0.2	0.1	0.1	0.3	0.5
4	131	351	0.3	0.4	0.2	0.1	0.7	0.8
5	139	622	0.6	0.8	0.4	0.2	1.2	1.5
6	145	991	1.0	1.2	0.7	0.3	2.1	2.2
7	150	1438	1.4	1.8	1.0	0.4	3.0	3.2
8	155	1968	1.9	2.5	1.3	0.6	4.2	4.2
9	160	2648	2.5	3.3	1.7	0.9	5.7	5.4
10	165	3427	3.3	4.3	2.2	1.1	7.5	6.9
11	169	4336	4.2	5.4	2.8	1.5	9.4	8.5
12	173	5431	5.2	6.7	3.6	2.0	11.7	10.4
13	177	6715	6.4	8.2	4.6	2.6	14.4	12.6
14	181	8350	8.0	10.0	5.9	3.3	17.9	15.3
15	184	9996	9.6	11.8	7.2	4.1	21.3	17.9
16	187	12111	11.6	14.2	8.8	5.3	25.2	21.3
17	191	14606	14.0	16.9	11.0	6.8	29.6	25.0
18	194	17382	16.7	19.8	13.3	8.4	34.5	29.2
19	197	20695	19.8	23.2	16.3	10.6	40.0	33.8
20	200	24467	23.5	27.0	19.6	13.4	45.8	38.7
21	203	28627	27.4	31.1	23.5	16.6	51.4	44.4
22	206	33259	31.9	35.6	27.9	20.4	57.2	50.1
23	209	38301	36.7	40.6	32.6	24.7	62.8	56.2
24	212	43896	42.1	46.0	37.9	30.0	68.1	62.6
25	215	49868	47.8	51.8	43.6	36.0	73.4	68.6
26	218	55987	53.7	57.6	49.4	42.5	78.3	73.9
27	221	62457	59.9	63.7	55.8	49.6	82.9	79.0
28	225	68976	66.1	69.9	62.1	57.0	86.9	83.7
29	229	75344	72.2	75.7	68.5	64.5	90.2	87.7
30	233	81351	78.0	81.2	74.5	71.6	93.2	91.1
31	237	86721	83.1	85.9	80.1	78.1	95.3	94.0
32	242	91507	87.7	90.1	85.2	84.1	97.0	96.0
33	250	95475	91.5	93.4	89.5	89.0	98.2	97.5
34	252	98628	94.5	96.0	93.0	93.0	98.8	98.5
35	258	100817	96.6	97.6	95.6	95.7	99.3	99.3
36	265	102349	98.1	98.7	97.4	97.7	99.7	99.6
37	272	103264	99.0	99.3	98.6	98.8	99.8	99.8
38	280	103792	99.5	99.7	99.3	99.4	99.9	99.9
39	288	104076	99.7	99.8	99.7	99.7	100.0	100.0
40	297	104229	99.9	99.9	99.8	99.9	100.0	100.0
41	300	104347	100.0	100.0	100.0	100.0	100.0	100.0

*All cumulative distributions include students scored on the full set of items.

Math Grade 5

Raw Score	Scale Score	All Students Cumulative* #	All Students Cumulative* %	Male Cumul. %	Female Cumul. %	White Cumul. %	Afr.-A. Cumul. %	Hisp. Cumul. %
0	100	30	0.0	0.0	0.0	0.0	0.1	0.1
4	108	81	0.1	0.1	0.1	0.0	0.2	0.2
5	119	188	0.2	0.2	0.1	0.1	0.4	0.4
6	128	399	0.4	0.5	0.3	0.2	0.9	0.7
7	136	736	0.7	0.9	0.5	0.3	1.8	1.4
8	143	1198	1.2	1.4	0.9	0.5	2.8	2.1
9	150	1815	1.8	2.2	1.4	0.7	4.2	3.3
10	156	2613	2.6	3.1	2.0	1.0	6.1	4.7
11	162	3603	3.5	4.1	2.9	1.5	8.3	6.4
12	168	4858	4.8	5.4	4.1	2.1	11.0	8.6
13	174	6419	6.3	7.0	5.6	2.9	14.2	11.3
14	179	8267	8.1	8.9	7.3	3.9	18.1	14.1
15	184	10466	10.3	11.0	9.5	5.2	22.3	17.6
16	189	13104	12.9	13.6	12.0	6.9	26.9	21.7
17	194	16110	15.8	16.5	15.0	9.1	31.8	26.1
18	200	19682	19.3	20.0	18.5	11.8	37.2	31.2
19	204	23626	23.2	23.8	22.5	15.0	43.0	36.1
20	208	27970	27.4	28.0	26.8	18.8	48.6	41.8
21	213	32685	32.1	32.4	31.6	23.1	54.3	47.4
22	217	37823	37.1	37.3	36.8	28.0	60.0	53.2
23	222	43394	42.6	42.6	42.5	33.6	65.6	58.9
24	226	49055	48.1	48.0	48.2	39.6	70.9	64.2
25	230	54848	53.8	53.5	54.1	45.7	75.9	69.6
26	235	60620	59.5	59.1	59.8	52.1	80.3	74.8
27	239	66409	65.2	64.6	65.7	58.6	84.1	79.8
28	244	71919	70.6	69.9	71.2	64.8	87.7	84.0
29	250	77448	76.0	75.2	76.7	71.2	91.1	87.9
30	254	82520	81.0	80.1	81.8	77.2	93.7	91.1
31	259	87108	85.5	84.6	86.3	82.6	95.9	93.7
32	265	91063	89.3	88.6	90.1	87.4	97.2	95.9
33	271	94335	92.5	91.8	93.3	91.2	98.4	97.4
34	279	97121	95.3	94.7	95.9	94.5	99.1	98.6
35	287	99058	97.2	96.7	97.7	96.8	99.6	99.3
36	298	100493	98.6	98.3	98.8	98.4	99.8	99.7
37	300	101937	100.0	100.0	100.0	100.0	100.0	100.0

*All cumulative distributions include students scored on the full set of items.

Math Grade 6

Raw Score	Scale Score	All Students Cumulative* #	All Students Cumulative* %	Male Cumul. %	Female Cumul. %	White Cumul. %	Afr.-A. Cumul. %	Hisp. Cumul. %
0	100	10	0.0	0.0	0.0	0.0	0.0	0.0
2	109	32	0.0	0.0	0.0	0.0	0.1	0.1
3	123	88	0.1	0.1	0.1	0.0	0.2	0.1
4	134	199	0.2	0.2	0.2	0.1	0.5	0.3
5	142	447	0.4	0.5	0.4	0.2	1.2	0.7
6	150	916	0.9	1.0	0.7	0.3	2.3	1.5
7	156	1664	1.6	1.8	1.3	0.7	4.1	2.7
8	162	2796	2.7	3.1	2.3	1.2	6.8	4.3
9	167	4341	4.2	4.6	3.7	1.9	10.3	6.5
10	172	6377	6.2	6.7	5.5	3.0	15.0	9.4
11	177	8719	8.4	9.1	7.6	4.2	20.0	12.9
12	181	11612	11.2	12.0	10.3	5.9	25.9	17.0
13	185	14860	14.3	15.3	13.3	8.0	31.5	21.7
14	189	18057	17.4	18.3	16.4	10.3	36.8	26.3
15	193	21750	21.0	21.9	20.0	13.0	42.6	31.1
16	200	25846	24.9	25.8	23.9	16.2	48.8	36.2
17	201	29999	28.9	29.7	28.1	19.6	54.3	41.4
18	204	34413	33.2	33.9	32.4	23.5	59.5	46.9
19	208	38883	37.5	38.0	37.0	27.6	64.4	52.1
20	211	43515	42.0	42.2	41.7	32.1	68.9	57.3
21	215	48247	46.5	46.6	46.4	36.7	73.4	62.4
22	219	53047	51.2	51.0	51.3	41.6	77.4	67.2
23	222	57792	55.7	55.3	56.2	46.6	80.8	72.0
24	226	62573	60.4	59.9	60.8	51.8	84.2	76.3
25	229	67199	64.8	64.2	65.4	56.9	87.1	80.3
26	233	71631	69.1	68.4	69.7	62.0	89.6	83.6
27	237	75948	73.3	72.6	74.0	67.1	91.7	86.8
28	241	80089	77.2	76.6	77.9	72.0	93.5	89.6
29	245	84001	81.0	80.5	81.6	76.7	95.0	92.0
30	250	87594	84.5	84.0	85.0	81.1	96.3	94.1
31	254	91042	87.8	87.4	88.2	85.3	97.4	95.6
32	258	94167	90.8	90.5	91.2	89.1	98.2	97.0
33	263	96886	93.4	93.2	93.8	92.4	98.9	97.9
34	269	99229	95.7	95.4	96.0	95.1	99.4	98.8
35	276	101146	97.6	97.3	97.8	97.3	99.7	99.4
36	284	102524	98.9	98.7	9.09	98.8	99.9	99.7
37	296	103308	99.6	99.6	99.7	99.6	100.0	99.9
38	300	103688	100.0	100.0	100.0	100.0	100.0	100.0

*All cumulative distributions include students scored on the full set of items.

Math Grade 7

Raw Score	Scale Score	All Students Cumulative* #	All Students Cumulative* %	Male Cumul. %	Female Cumul. %	White Cumul. %	Afr.-A. Cumul. %	Hisp. Cumul. %
0	100	33	0.0	0.0	0.0	0.0	0.1	0.0
2	111	164	0.2	0.2	0.2	0.1	0.4	0.2
3	126	507	0.5	0.5	0.4	0.2	1.2	0.7
4	137	1240	1.2	1.3	1.0	0.5	2.9	1.9
5	146	2642	2.5	2.8	2.2	1.1	6.1	4.1
6	154	4813	4.6	5.0	4.1	2.2	10.9	7.4
7	161	7808	7.4	8.0	6.9	3.8	17.0	11.7
8	167	11485	11.0	11.6	10.2	5.8	24.2	17.3
9	172	15777	15.0	15.7	14.3	8.4	32.1	23.6
10	178	20451	19.5	20.2	18.7	11.5	39.9	30.0
11	183	25309	24.1	24.7	23.4	15.0	47.1	36.7
12	188	30171	28.8	29.2	28.3	18.9	53.5	43.0
13	192	35326	33.7	33.9	33.4	23.3	59.6	49.4
14	200	40404	38.5	38.4	38.6	27.8	65.2	55.4
15	201	45367	43.3	42.9	43.6	32.5	70.2	60.6
16	205	50199	47.9	47.3	48.4	37.3	74.5	65.6
17	209	54917	52.4	51.5	53.2	42.1	78.3	70.3
18	213	59465	56.7	55.5	57.9	46.8	81.9	74.5
19	217	63663	60.7	59.5	62.0	51.4	84.7	78.1
20	221	67724	64.6	63.2	66.0	55.8	87.4	81.4
21	225	71690	68.3	66.8	69.9	60.3	89.5	84.5
22	228	75386	71.9	70.3	73.5	64.7	91.4	86.9
23	232	78968	75.3	73.7	77.0	68.9	93.1	89.2
24	236	82224	78.4	76.8	80.1	72.7	94.5	91.3
25	240	85346	81.4	79.8	82.9	76.5	95.7	93.1
26	244	88278	84.2	82.7	85.7	80.1	96.7	94.6
27	250	91027	86.8	85.4	88.3	83.5	97.6	95.8
28	253	93631	89.3	87.9	90.7	86.7	98.1	96.8
29	257	95926	91.5	90.3	92.7	89.6	98.7	97.7
30	262	98020	93.4	92.5	94.5	92.1	99.2	98.5
31	268	99927	95.3	94.4	96.1	94.4	99.5	99.1
32	274	101479	96.7	96.1	97.4	96.3	99.7	99.5
33	281	102675	97.9	97.4	98.4	97.7	99.8	99.7
34	289	103533	98.7	98.4	99.0	98.7	99.9	99.9
35	297	104131	99.3	99.1	99.5	99.3	100.0	99.9
36	300	104896	100.0	100.0	100.0	100.0	100.0	100.0

*All cumulative distributions include students scored on the full set of items.