

Grades 3 and 4
New Jersey Assessment of Skills
and Knowledge

TECHNICAL REPORT

MARCH 2004

PTM 1505.11
Copyright © 2004 by New Jersey Department of Education
All rights reserved

TABLE OF CONTENTS

PART 1: INTRODUCTION	1
1.1 Description of the New Jersey Assessment of Skills and Knowledge (NJ ASK)	1
1.2 State-Level Results	2
1.3 NJ ASK Organizational Support.....	3
PART 2: TEST DEVELOPMENT	4
2.1 Test Specifications	4
Language Arts Literacy	5
Mathematics.....	5
2.2 Development of Test Items	6
2.3 Item Review Process.....	8
2.4 Item Use.....	10
2.5 Test Forms Assembly	10
2.6 Quality Control for Test Construction.....	11
PART 3: TEST ADMINISTRATION	12
3.1 Participation	12
3.2 Test Security Procedures	13
3.3 Test Administration Procedures.....	13
3.4 Test Accommodations	14
PART 4: SCORING.....	15
4.1 Multiple Choice Items	15
4.2 Open Ended Items.....	15
Scorer Selection	16
Range Finding.....	16
Scorer Training	16
Scoring Procedures	16
4.3 Quality Control Procedures in Data Preparation	17
PART 5: STANDARD SETTING	17
5.1 Language Arts Literacy – Body of Work	18
5.2 Mathematics – Item-Descriptor Matching.....	18
5.3 Summary of Results.....	19
PART 6: ITEM-LEVEL STATISTICS.....	20
6.1 Classical Item Statistics	20
6.2 Speededness	28
6.3 Intercorrelations	31
6.4 Item Bias Statistics.....	35
PART 7: SCALING AND EQUATING	36
7.1 Scale Scores	36
7.2 Equating Language Arts Literacy	37
7.3 Equating Mathematics	40

PART 8: TEST STATISTICS	42
8.1 Summary Statistics	42
8.2 Classical Reliability Estimates of the Test Scores.....	49
8.3 Reliability of Performance Classifications	53
8.4 Conditional Estimate of Error at Each Cut-Score.....	55
8.5 Rater Reliability	56
Part 9: Validity	58
Content and Curricular Validity	58
Part 10: Reporting	59
10.1 Cycle I Reports	59
10.2 Cycle II Reports	63
10.3 State Summary Reporting.....	66
10.4 Interpreting Reports	67
10.5 Quality Control in Reporting	67
APPENDIX A: Statewide Cycle II Executive Summary Results	68
APPENDIX B: Additional Statewide Cycle II Results.....	78
APPENDIX C: Raw to Scale Score Conversions.....	95
REFERENCES	100

Tables and Figures

TABLE 1.2.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Total Student Group Tested –Grade 3	3
TABLE 1.2.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Total Student Group Tested –Grade 4	3
TABLE 2.1.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Total Points Possible by Content Area – Grades 3 & 4	6
Figure 2.3.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Item Approval Before Field Test.....	8
Figure 2.3.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Item Approval Before Operational Base Test.....	9
TABLE 2.3.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Number of Items Field Tested.....	10
TABLE 5.3.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Grade 3 Standard-Setting Results Recommended Cut Scores at the End of Phase 1 and Phase 2	19
TABLE 5.3.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Grade 3 Standard-Setting Results Recommended Cut Scores After Phase 2.....	20
TABLE 6.1.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Item Difficulty and Discrimination Summary Statistics for Dichotomously Scored and Open-Ended Items by Test Section and Cluster – Grade 3.....	22
TABLE 6.1.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Item Difficulty and Discrimination Summary Statistics for Dichotomously Scored and Open-Ended Items by Test Section and Cluster – Grade 4.....	23
TABLE 6.1.3 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Frequency Distributions of Item Difficulty Values and Biserial Discrimination Indices by Content Cluster Language Arts Literacy – Grade 3	24
TABLE 6.1.4 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Frequency Distributions of Item Difficulty Values and Biserial Discrimination Indices by Content Cluster Language Arts Literacy – Grade 4	25
TABLE 6.1.5 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Frequency Distributions of Item Difficulty Values and Biserial Discrimination Indices by Content Cluster Mathematics – Grade 3.....	26
TABLE 6.1.6 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Frequency Distributions of Item Difficulty Values and Biserial Discrimination Indices by Content Cluster Mathematics – Grade 4.....	27
TABLE 6.2.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Percentage of Students Omitting the Last Items of Each Test Part – Grade 3	29
TABLE 6.2.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Percentage of Students Omitting the Last Items of Each Test Part – Grade 4	30
TABLE 6.3.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Intercorrelations Among Major Content Clusters and Item Types – Grade 3	31
TABLE 6.3.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Intercorrelations Among Major Content Clusters and Item Types – Grade 4.....	32
TABLE 6.3.3 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Intercorrelations Among Content Areas and Clusters – Grade 3	33

TABLE 6.3.4 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Intercorrelations Among Content Areas and Clusters – Grade 4	34
TABLE 6.4.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) DIF Categories	36
TABLE 7.1.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Scaling Parameters for Base Forms	37
Figure 7.2.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) LAL Backwards and Forward Equating Designs	38
TABLE 7.2.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Language Arts Literacy Item Parameters – Grade 4	39
TABLE 7.2.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Forward Language Arts Literacy Fixed OE Item Step Parameters – Grade 4	39
TABLE 7.3.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Mathematics Item Parameters – Grade 4.....	41
TABLE 7.3.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Mathematics Fixed OE Item Step Parameters – Grade 4	42
TABLE 8.1.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students’ Raw Scores by Test Section – Grade 3	43
TABLE 8.1.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students’ Raw Scores by Test Section – Grade 4.....	43
TABLE 8.1.3 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students' Raw Scores and Percent Correct by Content Area – Grade 3.....	44
TABLE 8.1.4 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students' Raw Scores and Percent Correct by Content Area – Grade 4.....	45
TABLE 8.1.5 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students' Raw Scores and Percent Correct on the Dichotomously Scored Items by Content Area – Grade 3	46
TABLE 8.1.6 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students' Raw Scores and Percent Correct on the Dichotomously Scored Items by Content Area –Grade 4	47
TABLE 8.1.7 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students' Raw Scores and Percent Correct on the Open-Ended Items by Content Areas and Clusters – Grade 3	48
TABLE 8.1.8 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students' Raw Scores and Percent Correct on the Open-Ended Items by Content Areas and Clusters – Grade 4.....	49
TABLE 8.2.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability Estimates and Standard Errors of Measurement (SEM) for Content Areas and Clusters – Grade 3.....	50
TABLE 8.2.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability Estimates and Standard Errors of Measurement (SEM) for Content Areas and Clusters – Grade 4.....	51

TABLE 8.2.3 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability Estimates and Standard Errors of Measurement (SEM) for Dichotomously Scored Items Within Content Clusters – Grade 3	52
TABLE 8.2.4 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability Estimates and Standard Errors of Measurement (SEM) for Dichotomously Scored Items Within Content Clusters – Grade 4	53
TABLE 8.3.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability of Classification and Decision Accuracy – Grade 3	54
TABLE 8.3.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability of Classification and Decision Accuracy – Grade 4	55
TABLE 8.4.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Conditional Estimate of Error at Each Cut-Score.....	56
TABLE 8.5.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Consistency Between Raters Scoring Writing Tasks and Open-Ended Items - Grade 3	57
TABLE 8.5.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Consistency Between Raters Scoring Writing Tasks and Open-Ended Items – Grade 4	58
Figure 10.1.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Individual Student Report (ISR) – Front.....	60
Figure 10.1.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) Individual Student Report (ISR) – Back	61
Figure 10.2.1 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) School Report - Performance by Demographic Groups.....	64
Figure 10.2.2 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) School Report - Cluster Score Means	65
TABLE A.3.1 STATEWIDE PERFORMANCE BY DEMOGRAPHIC GROUPS SPRING 2004 -- GRADE 3 Language Arts Literacy	71
TABLE A.3.2 STATEWIDE PERFORMANCE BY DEMOGRAPHIC GROUPS SPRING 2004 -- GRADE 3 Mathematics	72
TABLE A.4.1 STATEWIDE PERFORMANCE BY DEMOGRAPHIC GROUPS SPRING 2004 -- GRADE 4 Language Arts Literacy	76
TABLE A.4.2 STATEWIDE PERFORMANCE BY DEMOGRAPHIC GROUPS SPRING 2004 -- GRADE 4 Mathematics	77
TABLE B.3.1 NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE BY DISTRICT FACTOR GROUP LANGUAGE ARTS LITERACY SECTION – Grade 3	78
TABLE B.3.2 NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE BY DISTRICT FACTOR GROUP MATHEMATICS SECTION – Grade 3	81
TABLE B.3.3 NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE LANGUAGE ARTS LITERACY AND MATHEMATICS FOR THE SPECIAL NEEDS DISTRICTS AS COMPARED TO ALL OTHER DISTRICTS – Grade 3.....	84
TABLE B.4.1 NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE BY DISTRICT FACTOR GROUP LANGUAGE ARTS LITERACY SECTION – Grade 4	85

TABLE B.4.2 NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE BY DISTRICT FACTOR GROUP MATHEMATICS SECTION – Grade 4	88
TABLE B.4.3 NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE LANGUAGE ARTS LITERACY AND MATHEMATICS FOR THE SPECIAL NEEDS DISTRICTS AS COMPARED TO ALL OTHER DISTRICTS – Grade 4.....	91
TABLE C.3.1 Conversion 2004 NJ ASK Language Arts Literacy – Grade 3	96
TABLE C.3.2 Conversion 2004 NJ ASK Mathematics – Grade 3.....	97
TABLE C.4.1 Conversion 2004 NJ ASK Language Arts Literacy – Grade 4	98
TABLE C.4.2 Conversion 2004 NJ ASK Mathematics – Grade 4.....	99

PART 1: INTRODUCTION

The purpose of this Technical Report is to provide information about the New Jersey Assessment of Skills and Knowledge (NJ ASK) administered as an operational assessment in March 2004. This report is intended for use by those who evaluate tests, interpret scores, or use test results for making educational decisions. It includes the following sections: test development, test administration, scoring, standard setting, item level statistics, scaling and equating, test statistics, validity, and score reporting. It includes references to additional reports and documents available for the NJ ASK.

1.1 Description of the New Jersey Assessment of Skills and Knowledge (NJ ASK)

The spring 2004 New Jersey Assessment of Knowledge and Skills (NJ ASK) was administered to students in grades three and four. It consisted of two content areas: Language Arts Literacy and Mathematics. Science was administered as a field-test to grade four students in 2004. Thus, science results will not be presented here. 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 Core Curriculum Content Standards. 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 Core Curriculum Content Standards in kindergarten through grade four. 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 scores are reported as scale scores and performance levels in each of the content areas. Following are the score ranges and their associated performance level.

- 100-199 Partially Proficient
- 200-249 Proficient
- 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 in need of instructional support.

The NJ ASK was administered between March 16 and March 19, 2004. The Language Arts Literacy and Mathematics tests were administered to 104,962 total students in grade 3. Performance levels for the grade 3 NJ ASK tests were established by panels of educators during sessions held between June 28 and July 6, 2004 and performance standards were approved by the New Jersey State Board of Education on July 7, 2004. The grade 4 performance standards for Mathematics were set in 1999, and the standards for grade 4 Language Arts Literacy were established in 2001. The Language Arts Literacy and Mathematics tests were administered to 105,340 total students in grade 4.

1.2 State-Level Results

This section includes two tables summarizing statewide test results for the 2004 administration of the NJ ASK. Tables 1.2.1 and 1.2.2 show the number and percentage of students in each performance category (i.e., Partially Proficient, Proficient, and Advanced Proficient) and the mean scale score for all students in Language Arts Literacy, and Mathematics in grades 3 and 4, respectively. The “number of students tested” is based on all students who received a test booklet, excluding those who were voided or APA exempt with no scale scores.

NOTE: Percentages shown in tables through this *Technical Report* may not total 100 due to rounding.

Following is a list of four state-level highlights for all students.

- Of the 103,414 grade 3 students with valid scale scores in Language Arts Literacy in Spring 2004, 20.7% scored in Partially Proficient; 75.6% scored in Proficient and 3.8% scored in Advanced Proficient (Table 1.2.1).
- Of the 103,559 grade 3 students with valid scale scores in Mathematics in Spring 2004, 23.4% scored in Partially Proficient; 53.8% scored in Proficient and 22.8% scored in Advanced Proficient (Table 1.2.1).
- Of the 103,818 grade 4 students with valid scale scores in Language Arts Literacy in Spring 2004, 17.9% scored in Partially Proficient; 77.5% scored in Proficient and 4.7% scored in Advanced Proficient (Table 1.2.2).
- Of the 103,770 grade 4 students with valid scale scores in Mathematics in Spring 2004, 27.9% scored in Partially Proficient; 46.4% scored in Proficient and 25.7% scored in Advanced Proficient (Table 1.2.2).

TABLE 1.2.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Total Student Group Tested –Grade 3**

TEST SECTION	NUMBER ^a OF VALID SCALE SCORES	PROFICIENCY LEVELS						SCALE SCORE MEAN
		PARTIALLY PROFICIENT (100-199)		PROFICIENT (200-249)		ADVANCED PROFICIENT (250-300)		
		No	%	No	%	No	%	
LANGUAGE ARTS LITERACY 2004	103,414	21,384	20.7%	78,135	75.6%	3,895	3.8%	215.5
MATHEMATICS 2004	103,559	24,260	23.4%	55,674	53.8%	23,625	22.8%	222.2

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID, AND APA EXEMPT WITH NO SCALED SCORES.

TABLE 1.2.2

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Total Student Group Tested –Grade 4**

TEST SECTION	NUMBER ^a OF VALID SCALE SCORES	PROFICIENCY LEVELS						SCALE SCORE MEAN
		PARTIALLY PROFICIENT (100-199)		PROFICIENT (200-249)		ADVANCED PROFICIENT (250-300)		
		No	%	No	%	No	%	
LANGUAGE ARTS LITERACY 2004	103,818	18,546	17.9	80,420	77.5	4,852	4.7	217.7
MATHEMATICS 2004	103,770	28,950	27.9	48,135	46.4	26,685	25.7	221.4

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID, AND APA EXEMPT WITH NO SCALED SCORES.

1.3 NJ ASK Organizational Support

The NJ ASK is administered by the Office of Evaluation and Assessment within the Department of Education. The staff of the Office of Evaluation and Assessment directs the implementation of the statewide assessment programs. In addition to planning, scheduling, and directing all NJ

ASK activities, the staff is extensively involved in numerous test review, security, and quality control procedures.

In 2003, the contract for developing and administering the NJ ASK was awarded to Educational Testing Service (ETS). ETS is the primary contractor working in partnership with Pearson Educational Measurement (PEM), The Grow Network, and Riverside Publishing Company. The major ETS activities include program management, test development, publication development and printing, supporting regional workshops that inform district test coordinators about the NJ ASK program, and psychometric support. Riverside Publishing Company develops the test items and supports the item review workshops. The major activities by PEM include: printing test books; distributing assessment materials in a secure manner; receiving, scanning, editing and scoring the answer documents; packaging, transporting and scoring open-ended responses; and providing data for score reporting. The Grow Network is responsible for producing, printing and shipping reports of test results to New Jersey pupils, parents/guardians, schools, districts and the state.

PART 2: TEST DEVELOPMENT

The Elementary School Proficiency Assessment (ESPA) was first administered at grade 4 from 1997 through 2002 to provide an early indication of student progress toward achieving the knowledge and skills identified in the Core Curriculum Content Standards (CCCS). ESPA was replaced in spring 2003 with the New Jersey Assessment of Skills and Knowledge (NJ ASK), a comprehensive, multi-grade assessment program. The purpose of these assessments is to provide indicators of student progress and to identify students who need additional instructional support in order to reach the CCCS. Details of the NJ ASK test development process are presented in this section.

2.1 Test Specifications

During the summer of 1996, three content committees consisting of 46 New Jersey educators developed the Elementary School Proficiency Assessment Content Domain Outline (February 1997), and a directory of test specifications and sample items for each content area to provide content/skill outlines and sample items. These directories describe the test, format of the items, and the scores to be generated by the test. This test specification work done by New Jersey educators serves as the foundation for all test item development.

The committees of New Jersey educators rely upon their expertise and the Core Curriculum Content Standards to design a test that is universally accessible to all grade 3 and grade 4 students and is composed of test questions that are age- and grade-appropriate. The material in the directories of test specifications and sample items as well as the Elementary School Proficiency Assessment Content Domain Outline is designed for use by curriculum specialists and teachers to improve instruction at the district, school and classroom levels.

In 2003, the ESPA became the NJ ASK. The NJ ASK is designed to measure the same Core Curriculum Content Standards as the ESPA. The items and test format of the NJ ASK are similar to those of the ESPA. In addition, the scale scores obtained from the NJ ASK are equivalent to those obtained from the ESPA. One difference between the two tests is the number of Mathematics clusters. In 2003, the Measurement and Geometry clusters of the ESPA were merged into one cluster for the NJ ASK. 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' achievements in reading and writing. Students read passages selected from published books, newspapers, and magazines as well as everyday text, and respond to related multiple-choice and open-ended questions.

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
 - Writing about Pictures
 - Writing About Poems
- Reading
 - Working with Text
 - Analyzing Text

For an in depth description of the NJ ASK Language Arts Literacy Test Specifications visit the NJ Department of Education website at:

<http://www.njpep.org/assessment/TestSpecs/LangArts/AssessOverview.html#CONTENT>

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 cluster in Mathematics contains one open-ended item. For an in-depth description of the NJ ASK Mathematics Test Specifications visit the NJ Department of Education website at: <http://www.njep.org/assessment/TestSpecs/MathNJASK/index.html>

Table 2.1.1 summarize the total points possible for each of the content areas of the operational NJ ASK administered in March 2004 for grades 3 and 4.

TABLE 2.1.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Total Points Possible by Content Area – Grades 3 & 4**

Language Arts Literacy	Grade 3	Grade 4
Total	40 points	43 points
Writing	20 points	20 points
Writing/Picture	10 points	10 points
Writing/Poem	10 points	10 points
-----	-----	-----
Reading	20 points	23 points
Working with Text	9 points	6 points
Analyzing Text	11 points	17 points
Mathematics	Grade 3	Grade 4
Total	33 points	43 points
C4.1 - Number Sense & Numerical Operations	9 points	13 points
C4.2 - Geometry & Measurement	8 points	10 points
C4.3 - Patterns & Algebra	8 points	10 points
C4.4 - Data Analysis, Probability & Discrete Math	8 points	10 points
-----	-----	-----
Problem Solving	25.5 points	33 points

* Within a content area, cluster-level results show how students perform on the sets of items that measure particular knowledge and skills (clusters above the dotted line) or particular processes (clusters below the dotted line). Though an item on the NJ ASK can contribute to a cluster above the line (for example, Reading) as well as a cluster below the line (for example, Working with Text), each item is counted only once in the total score.

2.2 Development of Test Items

The March 2004 NJ ASK consists of two types of items:

1. Operational or base test items used to determine students' scores and
2. Field-test items evaluated for use as future base test items.

A team of Riverside Publishing Company subject area specialists and consulting item writers begin the NJ ASK item development process. These writers are teachers or former teachers who have a great deal of specialized knowledge concerning their area of content expertise. All item writers for the NJ ASK program have (1) previously written items for a professional test development company or (2) attended an item-writer training workshop held by Riverside.

The following steps outline the item development process:

1. NJDOE and Riverside: Create test and item specifications
2. Riverside: Select and train item writers
3. Item Writers: Write test items
4. Riverside: Conduct initial item review
5. Riverside: Conduct item review by experienced senior staff
6. NJDOE: Conduct content and bias review
7. Items are field tested.
8. NJDOE: Conduct Statistical Item Review
9. Approved items go into the item bank

The Riverside Publishing Company item development process for each testing cycle begins with a formal review of the Core Curriculum Standards and the item specifications. The NJ ASK Item Specifications detail the standards to be measured, the number of items to be written, the item formats to be used, and other specific directions for developing the items. All NJ ASK items must be written to measure the New Jersey Core Curriculum Content Standards.

Item-writer training sessions are convened by content area at the Riverside headquarters in Itasca, Illinois. The respective test development specialist for each content area conducts the training session. Training consists of a full-day session with the first-half day used for specific training in understanding the Core Curriculum Content Standards and the test specifications. The second half-day is used for practice item writing. At the training, each consulting item writer is asked to sign a Letter of Agreement. This letter specifies the confidentiality and security regulations. This agreement also outlines the ownership regulations. No confidential materials related to the project are released without explicit approval by the Office of Evaluation and Assessment in New Jersey Department of Education (NJDOE).

During the training, each item writer is given an item writer's manual that includes the following:

- An overview of the New Jersey Assessment of Skills and Knowledge
- A final test blueprint for each subject area and item specifications
- A description of the item formats to be used, including important characteristics of each format
- A description of the item writing process and measures to avoid writing biased items
- A listing of the security procedures to be followed during the item development process

All items written by item writers are reviewed, revised, and edited by Riverside subject area specialists and editors prior to review by the New Jersey Test Committees. Before any item is included on a field test or operational base test, it must have the approval of the committees, as well as the NJDOE.

As items are developed, Riverside documents each item's relevancy to the Core Curriculum Content Standards and the directories of test specifications. During this process, each item is assigned a unique item identification number. The number is used to track the item throughout the development process and later in the item bank.

2.3 Item Review Process

Once test items have been through initial item review and item review by experienced senior staff at Riverside, the test materials are prepared for test committees' reviews. Before any item is included on a field test or operational base test, it must have the approval of the New Jersey Assessment Content and Sensitivity Review Committees. Typically, the committees consist of experienced educators, curriculum experts, and measurement specialists. Committee members also represent the diversity of the state in terms of ethnicity and geographic regions.

The New Jersey Test Committee members provide expert judgments as to the alignment of each test item with the Core Curriculum Content Standards and the content-specific test specifications. Committee members are selected based on their level of content area knowledge and number of years of teaching experience. Additionally, special care is taken to select members who are representative of the various districts and District Factor Groups (DFGs) within the State. Prior to field-testing, the Office of Evaluation and Assessment staff and the Language Arts Literacy, Mathematics, or Science Committees review all items. The Committees review each test item to determine if the item meets test specifications and addresses an appropriate level of difficulty. Committees also ensure that test questions are not offensive and do not reinforce negative stereotypes, and that test questions appropriately reflect multicultural society. Figure 2.3.1 presents a sample of the form that must be marked "Definitely Use" or "Revise and Use With Approval" during review committee meetings before an item is included on a field test.

Figure 2.3.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Item Approval Before Field Test**

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

Sensitivity Sign-off

Date

Content Chairperson's Signature

Date

All test items are field tested and reviewed again before they can be used as operational or base test items. The committees meet to review the item statistics. ETS calculates item means, response frequencies, biserial correlations (with base test total scores), and other descriptive statistics. Prior to the presentation of items and statistics to reviewers, the New Jersey

Department of Education defined boundaries within which item statistics should fall. In general, items with p-values below .30 or above 0.95 were considered to be usable only if a strong content argument could be made for their inclusion in the item bank. An item could be flagged for low or high p-value and/or low biserial correlation with base test total scores.

Also, for the statistical item review, the Mantel-Haenszel statistic is calculated to show whether or not students are responding to an item in a way that their overall ability (as measured by the base test) would lead us to expect. The statistic allows the committees to examine group membership (by ethnicity or by gender). The Mantel-Haenszel statistic is used for a classification determination of category A, B, or C. An item in Category A shows no or minor relationship between group membership and performance. Category B items show small to moderate relationship between membership and performance. Category C items show a substantial relationship between group membership and item performance and must be examined carefully by the committees to make sure these items are not biased.

Figure 2.3.2 presents a sample of the form that must be marked “Definitely Use” or “Revise and Use With Approval” during review committee meetings of the field-test statistics before an item is included on an operational base test.

Figure 2.3.2

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Item Approval Before Operational Base Test**

Sensitivity			Content		
*Comments			*Comments		
Sensitivity Issue	Yes	No	Appropriate Difficulty	Yes	No
If Yes, identify category and explain*				P-Value = 0.65	
Mantel-Haenszel Category C			Biserial = 0.42		
W-AA _____	W-H _____	M-F _____			
Definitely Use			Definitely Use		
Revise and Use With Approval			Revise and Use With Approval		
Revise and Resubmit			Revise and Resubmit		
Do Not Use*			Do Not Use*		

Sensitivity Sign-off

Date

Content Chairperson's Signature

Date

Table 2.3.1 show the number of field-test items presented during the March 2004 field-test administration. A sampling plan was developed that randomly assigned field-test forms to districts. To the extent possible, this plan insured that the student group taking each field-test form would be representative of the DFG distribution of the New Jersey districts.

TABLE 2.3.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Number of Items Field Tested**

		Multiple-Choice Items		Open-Ended Items		Writing Activities	
		Presented	Accepted	Presented	Accepted	Presented	Accepted
Grade 3	Language Arts	48	46	9	9	6	6
	Mathematics	115	96	13	1	--	--
		Presented	Accepted	Presented	Accepted	Presented	Accepted
Grade 4	Language Arts	44	38	16	14	4	4
	Mathematics	86	61	18	4	--	--
	Science	144	123	26	16	--	--

2.4 Item Use

All field-test items approved for use on an operational test form are moved into the item bank. Test development staff members choose from the available banked items when building an operational test form. A test item is used operationally one time, unless the item is used a second time as an anchor item in Mathematics. After operational use, items are retired. A small number of previously used items have been released for practice.

2.5 Test Forms Assembly

There are 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
- 1) **Determine forms design** – Each form consists of a set of operational items plus a set of variable items. The variable items provide opportunities for meeting equating needs and field-testing new items. The number of variable sections for each grade and subject is dependent upon the pool of items available for field-testing.
- 2) **Select items that meet content specifications** – Each content area measures subsets of items called clusters. In LAL the clusters include: Writing (Writing about Pictures and Writing about Poems), and Reading (Working with Text and Analyzing Text). Also, in grade 3 LAL,

there are items included on the test that measure Reading First. In Mathematics the clusters include: Number Sense and Numerical Operations; Geometry and Measurement; Patterns and Algebra; Data Analysis, Probability, and Discrete Mathematics. There is also a process cluster called Problem Solving. In Science the clusters include: Life Science, Physical Science and Earth Science. Science also has a process cluster called Application. Test forms must be similar to previous NJ ASK forms in terms of the number of items, the number of points, and the distribution of the content.

- 3) **Evaluate statistical specifications** – As forms are created it is necessary to determine if the statistical specifications have been met. Statistical specifications based on previous forms provide guidelines for building new test forms. Spreadsheets (form matrices) are used to provide information on the statistical properties of newly created forms. These matrices contain the following statistics: Average p-value, biserial correlation and average IRT difficulty (among other statistics). These data are reviewed to make certain that current forms are not substantially harder or easier than previous forms. Linking designs are also evaluated at this stage.
- 4) **Final approval of forms** – Once the content and statistical specifications have been met for each grade and subject, the forms are approved by the ETS Statistical Coordinator and by the NJ DOE. The forms are then released for production and editorial reviews.

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

2.6 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 NJ DOE.
- 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.
- Verify and document items needing manipulative sheets (math only).
- Ensure that the test meets the psychometric specifications.
- Verify match of statistical data on item card to statistical data on test matrix.
- 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 equal or nearly equal distribution of answer choices for MC items.
- ❑ Verify correct answer key for each item.
- ❑ Content review of form by senior staff.
- ❑ Statistical review of form by Measurement Statistician.
- ❑ Send form to NJ DOE for review and approval.

PART 3: TEST ADMINISTRATION

The Spring 2004 New Jersey Assessment of Skills and Knowledge (NJ ASK) included Grade 3 and Grade 4 testing sections in Language Arts Literacy and Mathematics and a Science field test for Grade 4 students. The Language Arts Literacy section consists of reading passages, multiple-choice items, open-ended items, and writing tasks. The Language Arts Literacy section is administered over two days for both grades. The Mathematics section consists of multiple-choice and open-ended items that must be answered with the use of a calculator, and multiple-choice items that must be answered without the use of a calculator. The Mathematics section is administered over a two-day period for Grade 4 and a one-day period for Grade 3.

Language Arts Literacy and Mathematics field-test items are embedded within the sections of the regular test. The Science field test, which consisted of multiple-choice and open-ended items, was administered on one day. The make-up tests are scheduled by school districts for administration any morning during the week following the regular NJ ASK administration. Districts have the flexibility to choose which subjects are tested on which days of the make-up period.

3.1 Participation

General Education Students

The NJ ASK must be administered to all third- and fourth-grade students in New Jersey public schools except those whose Individual Education Program exempts them from taking the NJ ASK.

Limited English Proficient Students

Limited English Proficient (LEP) students must take the test according to federal guidelines for No Child Left Behind.

Students with Disabilities

Students with Disabilities in the third- and fourth-grade eligible for special education under the Individuals with Disabilities Education Act or eligible under Section 504 of the Rehabilitation Act of 1973 must take each subject area of the NJ ASK unless their Individualized Education Program (IEP) or 504 plan specifically states that they will not participate in one or more subject

areas of the test. Students who are ungraded must take the NJ ASK in the calendar year in which they are 9, 10, or 11 years old and when they are first instructed in the knowledge and skills tested. Students whose IEP exempts them from participation in the NJ ASK must participate in the Alternate Proficiency Assessment (APA).

3.2 Test Security Procedures

Standard Security Procedures

The NJ ASK test booklets and their contents are secure materials. Detailed procedures for maintaining the security of test materials while test materials are in the districts are outlined in the Test Administration Manual. It is the responsibility of school districts to guarantee the security of the test materials. Examiners, proctors, and other school personnel are prohibited from copying, reading, discussing, or disclosing any test items before, during, or after the test administration. When not being used during a test period, test materials are stored in a secure, locked place that is accessible only to individuals whose access is authorized by the school test coordinator. Inventory forms track test materials as they move from one location to another within the districts.

Security Breach Procedures

Breach test forms and examiner manuals are prepared in the event of a security breach. If the New Jersey Department of Education (NJ DOE) identifies a security breach during the test administration window the sub-contractor immediately removes the NJ ASK test materials from the involved district or school. The test books for the subject area affected are coded with a void code 5 indicating a security breach. If time permits (determined by NJ DOE) breach forms are delivered to the districts and districts are required to test the affected students in the subject area impacted. When students are re-tested during the test administration window scores are reported based on the breach form test scores. If a security breach is identified after the test administration window the impacted test books are coded void code 5 (security breach) and no test results are reported for that subject area. Students receive a score for the subject area that was not impacted by the security breach.

3.3 Test Administration Procedures

School test coordinators, examiners and proctors are responsible for the administration of the exam. Their responsibilities include

- distributing test materials each morning of testing,
- overseeing the recording on School Security Checklists of the transfer of test booklets,
- supervising testing, ensuring proper test administration procedures are followed according to the instructions in the provided Examiner Manuals,
- ensuring that accommodations/modifications listed in the IEPs/504 plans of students with disabilities are implemented

- monitoring any potential circumstances that may seriously interrupt or interfere with the test administration
- reporting any testing irregularities that occur during the administration
- notifying district test coordinator immediately of any missing test booklets
- scheduling make-up testing for any students who missed one or more days of the regular testing period.
- returning testing materials to contractors

3.4 Test Accommodations

General Education Students

General education students receive no special testing accommodations other than the standard room setup and materials distribution described in the Examiner Manual.

Accommodations and Modifications for Students with Disabilities

To ensure that students are tested under appropriate conditions, the Department of Education has adopted test accommodations and modifications that may be used when testing special populations of students. The content of the test typically remains the same, but administration procedures, setting, and answer modes may be adapted. Students requiring accommodations must be tested in a separate location from general education students.

Special education students must take the NJ ASK unless their IEP specifically exempts them. A student whose IEP exempts them from taking the NJ ASK must participate in the APA. Special education students may be tested using accommodations/modifications specified in the students' Individualized Education Programs (IEPs) that are approved by the Office of Evaluation and Assessment. Students who have a disability and are eligible under Section 504 of the Rehabilitation Act of 1973 may be tested using accommodations/modifications specified in the student's 504 plan that are approved by the Office of Evaluation and Assessment.

Large-print and Braille materials are provided to districts as required. Students completing a Braille version of the Mathematics section are instructed to bring a Braille ruler to the test session as well as a talking calculator. Students completing a large-print version of the test may use a ruler that is used during class instruction.

Students using the Braille test booklets are permitted to dictate their answers for multiple-choice questions to the examiner. Students taking the Braille test are also permitted to dictate their responses to the open-ended questions and all writing tasks. If dictation is used, the student is required to indicate all punctuation and must spell all key words.

Students using the large-print test booklets mark their answers for multiple-choice questions in the large-print version of the test booklet. Visually impaired students may use special equipment such as a typewriter or computer, if appropriate, for the open-ended questions and writing tasks. For 2004, the Braille versions differed from the standard versions of the tests as some items were

omitted. These items are noted in the student's copy of the test. A list is provided to the examiners along with the supplemental instructions for administering the large-print and Braille versions of the test.

Accommodations for Limited English Proficient Students

NCLB prohibits exemptions from testing based on LEP status. However, limited English proficient (LEP) students were tested with one or more accommodations in the test administration procedures. Permitted accommodations include the following:

- additional time up to 150% of the administration times indicated
- translation of the test directions only into the student's native language (translations of passages, items, prompts, and tasks are NOT permitted)
- use of a bilingual dictionary

Students who received translated test directions were tested in a location separate from students tested with directions read in English only.

PART 4: SCORING

4.1 Multiple Choice Items

Before any documents are scanned, a complete check of the scanning system is conducted. A mock set of answer documents are gridded to cover all response ranges, demographic data, blanks, double grids and other responses. Mock student records are created to verify that each gridding possibility is processed correctly by the scanning program. The output file that is created is thoroughly hand-checked against each answer document after each stage to ensure that the scanners are capturing all marks correctly. When the program output is confirmed to match the expected results, a formal sign-off process takes place.

The scoring keys are reviewed and approved prior to entry into the scoring system, and once entered, are verified. The multiple-choice scoring process entails multiple reviews for accuracy performed by independent staff on each key in every form.

4.2 Open Ended Items

Scoring of Open-Ended (OE) items involves having trained scorers read each student response by at least two readers. The student responses are assigned points by the scorers based on rules outlined in scoring rubrics. For more information about the scoring rubrics, readers are referred to the Cycle I Interpretation Manual.

Scorer Selection

The selection of scorers for the constructed response items is made from a large pool of candidates who meet stringent qualifications. Scorers must have, at a minimum, a four-year college degree, and must complete an individual interview. Preference is given to individuals with degrees and backgrounds related to language arts, mathematics and science, and experience in performance scoring. If appropriate, they are also asked to complete a grammar placement test and submit an original writing sample. Scoring supervisors are chosen based on subject area expertise, along with strong organizational abilities and communication skills. Scoring supervisors must demonstrate the ability to assist Scoring Directors in training, calibration and discussion sessions by successfully articulating the unique scoring criteria and their application.

Range Finding

Rangefinding sessions are conducted using a range of photocopied student responses for each item. These responses are used to expand and refine existing anchor sets (selected examples of student work representing the score points), to be used in the training for operational scoring.

Scorer Training

Comprehensive training for scorers is provided via an online training system. This system incorporates scoring guides, fully annotated sample responses, practice exercises and qualifying sets. The training is user-driven and interactive and scorers are able to set their own pace.

The scoring guides present the rubrics with descriptions of each score level, and guidelines are provided on how to properly apply the scoring criteria. Annotated papers are chosen to clearly represent each designated score point. These student responses serve as the primary points of reference for scorers as they internalize the rubric during training. All scorers have access to this anchor set whenever they are scoring, and are directed to refer to it regularly.

Practice sets of student responses are used during training to help scorers become more experienced in applying the rubric. The use of these practice sets provides guidance to scorers in defining the line between score points and in applying the scoring criteria to a wider range of types of responses.

Sets of student responses which incorporate a range of student performance levels are used to confirm that the trainees can correctly assign the full range of scores. Candidates must demonstrate acceptable performance on these sets in order to qualify as a scorer.

Scoring Procedures

Once trained, the scorers review and score responses using an electronic scoring system, which is accessible from multiple locations. The security protocols within the system are designed to ensure the individual who received the training and is qualified to score is the individual who is scoring the responses. Scoring rate, reliability and validity statistics are monitored by the system

and by supervisors to manage scoring performance and to identify changes or trends in the scorer's performance. If a scoring anomaly is suspected, the problematic scorer can be locked from the system and all or a portion of their work may be reset to address a scoring quality issue. The system assigns priority to student responses within the pool of available student responses based on a first-in and first-out system, and delivers to the scorer the next eligible response from the pool. Items requiring second reads are given priority over unscored responses, and the system prevents a response from receiving the first and second scores from the same scorer.

All responses are scored by two scorers. If the first and second scores for a response are non-adjacent (e.g., one reader assigns a "5", and the second reader a "3"), the response will be forwarded to a scoring supervisor, who will review and score the response to resolve the discrepancy.

Qualified scorers are authorized to assign valid score points or the "Blank" condition code to responses. Supervisory staff score items sent to them for review, non-adjacent items requiring resolution and all other condition codes (No Response, Off Topic, Not English, Wrong Format, etc).

4.3 Quality Control Procedures in Data Preparation

All information gridded on the students' test booklets is automatically scanned and a series of edit checks are applied during and after the scanning process, prior to storage of the data in a master database. Some student demographic data in the database may be modified through an online password-protected system accessible to specified individuals within the districts.

The master database is the origination of all data for files and reports for the testing administration. This includes all paper reporting, reporting via CDs, and files for the preparation of other State reporting.

Each time data is extracted from the master database for any of the reporting cycles or other files required by the DOE, the extracted data is put through a series of quality control checks to ensure its accuracy for that reporting cycle or file. Once the extracted data has been verified as correct and complete, the reporting cycle continues with the production of reports or files.

PART 5: STANDARD SETTING

After the March 2004 administration, standard setting workshops were held in June for the grade 3 Language Arts Literacy and Mathematics tests. Standard setting was conducted in two phases. Phase 1 involved approximately 20 educators (per panel) from across the state of New Jersey meeting for 2½ to 4 days and using a research-based standard setting method to recommend cut scores for the Proficient and Advanced Proficient levels. Phase 2 immediately followed Phase 1 and involved 3 teachers from each of the two Phase 1 panels as well as 3 additional policymakers from the state. The Phase 2 panel reviewed the Phase 1 cut scores along with additional information about the percentage of students who would be classified in each

level. This additional information included the percentage of students in all reporting categories (e.g., economically disadvantaged) who would reach Proficient and Advanced Proficient and the percentages currently reaching those levels in grade 4. They then provided their recommendations for cut scores, which were presented to the New Jersey State Board of Education for review and adoption.

Two different methods were used to set standards because the two subject areas have different test specifications. Following is a brief summary of each procedure and the results. For more information about the standard-setting workshops a full report is available from the NJDOE.

5.1 Language Arts Literacy – Body of Work

Because the LAL test was comprised primarily of open-ended items and writing prompts, a holistic method was chosen to determine cut scores. The Body of Work method required panelists to review entire student booklets, including responses to both open-ended and multiple-choice items, and determine whether the skills and knowledge demonstrated in the booklet best match the performance level descriptors for Partially Proficient, Proficient, or Advanced. The panelists were not told what the scores were for each booklet, but the standard setting facilitators used the information on judges' ratings in combination with the scores for each booklet to calculate a cut score for each level. Body of Work was conducted over two rounds. In the first round rangefinding, panelists were given 30 booklets with scores ranging from 4 to 38 points out of 40. Based on the ratings of these 30 booklets, a second set of booklets were pulled for round 2, the pinpointing round. After the Rangefinding round, the preliminary cut scores were calculated to be 20 points for Proficient and 30 points for Advanced Proficient. Another 22 booklets were selected to cover the range of 15 to 25 points for the Partially Proficient/Proficient cut score and 22 more booklets at 25 to 35 points for the Proficient/Advanced Proficient cut score. These 44 new booklets were used in the Pinpointing round to determine exactly where the cut scores should fall within the initial ranges.

In both procedures, panelists received “consequence” information about the percentage of students who took the test in March that would be categorized as Partially Proficient, Proficient, or Advanced Proficient. The percentages given to the panelists were based on the cut scores set after the first Pinpointing rating in Body of Work and after Round 2 in ID Matching. They then had the chance to discuss this information with their peers and make final adjustments to their ratings before the conclusion of Phase 1.

5.2 Mathematics – Item-Descriptor Matching

For Mathematics, which consisted primarily of multiple-choice items, an item mapping procedure called Item-Descriptor (ID) Matching was used. This is a variant of the Bookmark procedure that orders the operational items by difficulty as determined by the scale location of the items. Thus, the items that students performed best on appear first in an ordered test booklet and the items they performed worst on appear last. Panelists first go through the ordered test booklet and match the knowledge and skills required by the each item to the knowledge and

skills listed in the performance level descriptors. That is, they ask themselves what one has to know and be able to do to answer an item correctly and then determine whether the knowledge and skills more closely match the descriptions of Partially Proficient, Proficient, or Advanced Proficient. Once they have matched each item to a performance level descriptor, they then determine the location of the cut score that best separates Partially Proficient performance from Proficient performance and Proficient from Advanced Proficient performance. ID Matching is conducted over three rounds with panelists receiving feedback about their ratings and having a chance to discuss their ratings with their peers between rounds.

5.3 Summary of Results

Overall, panelists’ judgments about the cut scores converged from one round to the next, showing strong agreement by the end of the final round of Phase 1 and even more agreement in Phase 2. Table 5.3.1 shows the recommended cut scores at the end of Round 2 (the equivalent of the first Pinpointing rating in Body of Work) before the panelists saw the consequences data, at the end of Round 3 and then after Phase 2. The standard error of measurement (SEM) shows the degree of uncertainty in a student’s score on the test around the cut score, and the standard error of judgment (SEJ) is related to the variance in panelists’ judgments around the cut score. Overall, the SEJs decreased across rounds, indicating converging opinions. Table 5.3.1 also shows that the Phase 2 panelists adopted the Phase 1 recommendations for three of the four cut scores and only modified the recommendation for the proficient cut score in mathematics by 2 points—within 1 SEM of the cut score recommended at the end of Phase 1.

TABLE 5.3.1

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

Grade 3 Standard-Setting Results

Recommended Cut Scores at the End of Phase 1 and Phase 2

	Phase 1 - Round 2		Phase 1 - Round 3		Phase 2	
	Proficient	Advanced Proficient	Proficient	Advanced Proficient	Proficient	Advanced Proficient
LAL						
Cut Score	19.5	32.5	18	30.5	18	30.5
SEM	2.5	2.0	2.5	2.0	2.5	2.0
SEJ	0.30	0.40	0.14	0.32	0.09	0.00
Math						
Cut Score	14.5	32.0	15	27.5	17.0	27.5
SEM	2.5	2.0	2.5	2.0	2.5	2.0
SEJ	0.50	0.31	0.32	0.35	0.25	0.12

Table 5.3.2 shows the final cut scores that were brought to the State Board of Education for their review and approval. The four cut scores recommended by the Phase 2 panel were presented

along with the consequences data showing the percentage of grade 3 students who would be categorized as Partially Proficient, Proficient, and Advanced Proficient. In addition to these tables, the State Board was also provided with consequence data for students in each reporting category, such as gender, race/ethnicity, and economic status. The State Board voted unanimously to adopt the recommended cut scores for the NJ ASK3.

TABLE 5.3.2

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Grade 3 Standard-Setting Results
Recommended Cut Scores After Phase 2**

	Cut score for Proficient	Cut Score for Advanced Proficient	% Partially Proficient	% Proficient	% Advanced Proficient
LAL	18.0	30.5	21.6%	74.6%	3.8%
Math	17.0	27.5	23.8%	53.4%	22.8%

PART 6: ITEM-LEVEL STATISTICS

6.1 Classical Item Statistics

For each administration, classical item analyses are completed prior to item calibration, scaling and equating. These statistics are calculated again once all of the data are available. These analyses involve computing, for every item in each form, a set of statistics based on classical test theory. Each statistic is designed to provide some key information about the quality of each item from an empirical perspective. The statistics estimated for the NJ ASK are described below.

- Classical item difficulty (“P-Value”):
This statistic indicates the percent of examinees in the sample that answered the item correctly. Desired p-values generally fall within the range of 0.25 to 0.90.
- Item discrimination (“r-biserial)¹:

¹ The estimated polyserial correlation between scores on the item and on the criterion is computed by the formula:

$$r_{polyreg} = \frac{\beta_i \sigma_x}{\sqrt{\beta_i^2 \sigma_x^2 + 1}}$$

This statistic is measured by the polyserial correlation between the item score and the test criterion score and describes the relationship between performance on the specific item and performance on the entire form. The higher the value, the better the task of separating the examinees. Items with negative correlations can indicate serious problems with the item content (e.g., multiple correct answers or unusually complex content), or can indicate that students have not been taught the content. For LAL, the test criterion score was the number-correct score on the MC items, plus the weighted CR item score. For mathematics, the test criterion score was the number-correct score.

- The proportion of students choosing each response option:
These statistics indicate the percent of examinees that select each of the available answer options and the percent of examinees that omitted the item.
- Distracter analyses for MC items.
The GENASYS system (GENASYS is a proprietary ETS item analysis software program) provides graphical displays of the data for each option, which are reviewed.
- Percent of students omitting an item:
This statistic is useful for identifying problems with test features such as testing time and item/test layout. Typically, we would expect that if students have an adequate amount of testing time, 95% of students should attempt to answer each question. When a pattern of omit percentages exceeds 5% for a series of items at the end of a timed section, this may indicate that there was insufficient time for students to complete all items. Alternatively, if the omit percentage is greater than 5% for a single item, this could be an indication of an item/test layout problem. For example, students might accidentally skip an item that follows a lengthy stem.

In Tables 6.1.1 and 6.1.2, summary statistics are given that describe the difficulty and discrimination of the items comprising each cluster for grades 3 and 4, respectively. For dichotomously scored items, means and standard deviations of proportion-correct values (p-values) and r-biserials are given. For the open-ended items, the index of item difficulty was calculated by dividing students' average scores on an item by the maximum possible score on the item. Item discrimination for each open-ended item is the correlation between students' item score and their total score on the test section. For both the item-test correlation and the r-biserial correlation, students' total test scores were expressed in terms of the raw score metric.

where the β_i are a series of parameters estimated by maximum likelihood from the item analysis data (Drasgow, 1988; Lewis & Thayer, 1996).

TABLE 6.1.1

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Item Difficulty and Discrimination Summary Statistics
for Dichotomously Scored and Open-Ended Items
by Test Section and Cluster – Grade 3**

NJ ASK Test Section/Cluster	Dichotomous			Open-Ended		
	Item Difficulty		Item Discrimination	Item Difficulty		Item Discrimination
	Mean	S.D.	Mean	Mean	S.D.	Mean
Language Arts Literacy	0.71	0.14	0.57	0.47	0.05	0.74
Writing	--	--	--	0.48	0.02	0.79
Writing/Picture	--	--	--	0.50	--	0.81
Writing/Poem	--	--	--	0.47	--	0.76
Reading	0.71	0.14	0.57	0.46	0.08	0.70
Working with Text	0.73	0.12	0.58	--	--	--
Analyzing Text	0.65	0.22	0.54	0.46	0.08	0.70
Mathematics	0.72	0.10	0.55	0.52	0.16	0.63
Number Sense & Numerical Operations	0.74	0.09	0.55	--	--	--
Geometry & Measurement	0.77	0.10	0.48	0.34	--	0.66
Patterns & Algebra	0.66	0.03	0.56	0.64	--	0.55
Data Analysis, Probability & Discrete Math	0.67	0.13	0.59	0.57	--	0.68
Problem Solving	0.70	0.09	0.56	0.52	0.16	0.63

TABLE 6.1.2

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Item Difficulty and Discrimination Summary Statistics
for Dichotomously Scored and Open-Ended Items
by Test Section and Cluster – Grade 4**

NJ ASK Test Section/Cluster	Dichotomous			Open-Ended		
	Item Difficulty		Item Discrimination	Item Difficulty		Item Discrimination
	Mean	S.D.	Mean	Mean	S.D.	Mean
Language Arts Literacy	0.70	0.13	0.52	0.50	0.05	0.76
Writing	--	--	--	0.54	0.02	0.81
Writing/Picture	--	--	--	0.56	--	0.82
Writing/Poem	--	--	--	0.53	--	0.79
Reading	0.70	0.13	0.52	0.47	0.03	0.73
Working with Text	0.73	0.11	0.53	--	--	--
Analyzing Text	0.66	0.16	0.51	0.47	0.03	0.73
Mathematics	0.67	0.15	0.53	0.47	0.15	0.70
Number Sense & Numerical Operations	0.75	0.14	0.54	0.60	0.12	0.67
Geometry & Measurement	0.68	0.16	0.52	0.29	--	0.68
Patterns & Algebra	0.58	0.13	0.52	0.47	--	0.79
Data Analysis, Probability & Discrete Math	0.61	0.14	0.53	0.41	--	0.70
Problem Solving	0.60	0.14	0.52	0.47	0.15	0.70

Frequency distributions of the March 2004 NJ ASK item p-values (difficulty values) and item discrimination indices are provided by content section and cluster for Language Arts Literacy and Mathematics in Tables 6.1.3, 6.1.4, 6.1.5 and 6.1.6. The top section of each table shows the distribution of item difficulty values; the bottom section shows the distribution of r-biserial indices.

TABLE 6.1.3**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Frequency Distributions of Item Difficulty Values and Biserial Discrimination Indices by Content Cluster****Language Arts Literacy – Grade 3**

Item Statistics	Working With Text	Analyzing Text	Read First	Total
ITEM DIFFICULTY: P-VALUES				
.800+	4	1	4	5
.700 - .799	1	0	1	1
.600 - .699	2	0	2	2
.500 - .599	2	1	1	3
<.500	0	1	0	1
MEAN P-VALUE	0.73	0.65	0.75	0.71
MEDIAN P-VALUE	0.76	0.56	0.79	0.73
ITEM DISCRIMINATION: BISERIAL CORRELATIONS				
.50+	8	2	8	10
.40 - .49	1	1	0	2
.30 - .39	0	0	0	0
MEAN POINT-BISERIAL	0.58	0.54	0.60	0.57
MEDIAN POINT-BISERIAL	0.61	0.54	0.61	0.58
TOTAL NUMBER OF ITEMS	9	3	8	12

TABLE 6.1.4**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Frequency Distributions of Item Difficulty Values and Biserial
Discrimination Indices by Content Cluster****Language Arts Literacy – Grade 4**

Item Statistics	Working With Text	Analyzing Text	Total
ITEM DIFFICULTY: P-VALUES			
.800 - .899	2	1	3
.700 - .799	2	1	3
.600 - .699	0	1	1
.500 - .599	2	1	3
<.500	0	1	1
MEAN P-VALUE	0.73	0.66	0.70
MEDIAN P-VALUE	0.76	0.67	0.70
ITEM DISCRIMINATION: BISERIAL CORRELATIONS			
.50+	3	3	6
.40 - .49	3	2	5
.30 - .39	0	0	0
MEAN POINT-BISERIAL	0.53	0.51	0.52
MEDIAN POINT-BISERIAL	0.52	0.55	0.55
TOTAL NUMBER OF ITEMS	6	5	11

TABLE 6.1.5

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Frequency Distributions of Item Difficulty Values
and Biserial Discrimination Indices by Content Cluster**

Mathematics – Grade 3

Item Statistics	Number Sense & Numerical Operations	Geometry & Measurement	Patterns & Algebra	Data Analysis, Probability & Discrete Math	Problem Solving	Total Test
ITEM DIFFICULTY: P-VALUES						
.900+	0	0	0	0	0	0
.800 - .899	3	3	0	0	2	6
.700 - .799	5	1	1	3	7	10
.600 - .699	3	1	4	1	7	9
.500 - .599	1	0	0	0	0	1
<.500	0	0	0	1	1	1
MEAN P-VALUE	0.74	0.77	0.66	0.67	0.70	0.72
MEDIAN P-VALUE	0.76	0.81	0.65	0.71	0.71	0.72
ITEM DISCRIMINATION: BISERIAL CORRELATIONS						
.50+	8	1	4	5	12	18
.40 - .49	4	4	1	0	5	9
.30 - .39	0	0	0	0	0	0
.20 - .29	0	0	0	0	0	0
MEAN POINT-BISERIAL	0.55	0.48	0.56	0.59	0.56	0.55
MEDIAN POINT-BISERIAL	0.52	0.47	0.56	0.58	0.56	0.54
Total Number of Items	12	5	5	5	17	27

TABLE 6.1.6

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Frequency Distributions of Item Difficulty Values
and Biserial Discrimination Indices by Content Cluster**

Mathematics – Grade 4

Item Statistics	Number Sense & Numerical Operations	Geometry & Measurement	Patterns & Algebra	Data Analysis, Probability & Discrete Math	Problem Solving	Total Test
ITEM DIFFICULTY: P-VALUES						
.900 - .999	2	0	0	0	0	2
.800 - .899	2	2	0	1	1	5
.700 - .799	4	2	1	1	4	8
.600 - .699	2	1	2	1	4	6
.500 - .599	0	1	3	2	5	6
<.500	1	1	1	2	4	5
MEAN P-VALUE	0.75	0.68	0.58	0.61	0.60	0.67
MEDIAN P-VALUE	0.75	0.74	0.55	0.58	0.61	0.69
ITEM DISCRIMINATION: BISERIAL CORRELATIONS						
.50 - .59	9	5	3	4	11	21
.40 - .49	2	2	4	2	6	10
.30 - .39	0	0	0	1	1	1
.20 - .29	0	0	0	0	0	0
MEAN POINT-BISERIAL	0.54	0.52	0.52	0.53	0.52	0.53
MEDIAN POINT-BISERIAL	0.54	0.51	0.49	0.51	0.51	0.52
Total Number of Items	11	7	7	7	18	32

6.2 Speededness

The NJ ASK is intended to provide sufficient time for all students to respond to almost all of the questions. The percentage of students omitting an item provides information about speededness, although it must be kept in mind that students can omit an item for reasons other than speededness (for example, choosing to not put effort into answering a constructed response item). Thus, if the percentage of omits is low, that implies that there is little speededness; if a percentage of omits is high, speededness, as well as other factors, can be the cause.

Tables 6.2.1 and 6.2.2 present data concerning the extent to which students omitted items. Table 6.2.1 shows that the percentage of grade 3 students omitting the Reading multiple-choice items was very small while the percentage of students omitting the Reading open-ended items varied from 1.2% to 1.9%. Table 6.2.1 also shows the percentage of grade 3 students omitting each of the last two Mathematics multiple-choice items in each part and all Mathematics open-ended items. The percentage of grade 3 students omitting the Mathematics multiple-choice items ranged from 0.3% to 2.1%. The percentage of grade 3 students omitting the Mathematics open-ended items ranged from 2.6% to 3.0%.

Table 6.2.2 shows that the percentage of grade 4 students omitting the Reading multiple-choice items was very small while the percentage of students omitting the Reading open-ended items varied from 0.5% to 3.2%. Table 6.2.2 also shows the percentage of grade 4 students omitting each of the last two Mathematics multiple-choice items in each part and all Mathematics open-ended items. The percentage of grade 4 students omitting the Mathematics multiple-choice items ranged from 0.7% to 4.4%. The percentage of grade 4 students omitting the Mathematics open-ended items ranged from 1.0% to 6.4%.

TABLE 6.2.1

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Percentage of Students Omitting the
Last Items of Each Test Part – Grade 3**

Test Section	Multiple - Choice		Open - Ended	
	Item Number	Percentage Omitting	Item Number	Percentage Omitting
Reading				
<u>First Part</u>	Item 5	0.3%		
	Item 6	0.4%	Item 7	1.2%
<u>Second Part</u>	Item 5	0.6%		
	Item 6	0.7%	Item 7	1.9%
Mathematics				
<u>Day 1</u>				
<u>First Part</u>	Item 2	0.3%		
	Item 3	0.7%		
<u>Second Part</u>	Item 5	1.4%		
	Item 6	1.4%		
<u>Third Part</u>	Item 12	1.1%		
	Item 13	1.6%	Item 14	2.7%
<u>Fourth Part</u>	Item 20	1.3%		
	Item 21	1.6%	Item 22	2.6%
<u>Fifth Part</u>	Item 28	0.6%		
	Item 29	2.1%	Item 30	3.0%

TABLE 6.2.2

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Percentage of Students Omitting the
Last Items of Each Test Part – Grade 4**

Test Section	Multiple – Choice		Open - Ended	
	Item Number	Percentage Omitting	Item Number	Percentage Omitting
Reading				
<u>First Part</u>	Item 4	0.2%	Item 6	0.5%
	Item 5	0.2%	Item 7	3.2%
<u>Second Part</u>	Item 5	0.3%		
	Item 6	0.3%	Item 7	0.8%
Mathematics				
<u>Day 1</u>				
<u>First Part</u>	Item 3	0.8%		
	Item 4	2.2%		
<u>Second Part</u>	Item 7	0.7%		
	Item 8	1.3%		
<u>Third Part</u>	Item 19	4.1%		
	Item 20	4.4%	Item 21	6.4%
<u>Fourth Part</u>	Item 26	0.9%	Item 28	3.2%
	Item 27	1.2%	Item 29	4.8%
<u>Day 2</u>				
<u>Fifth Part</u>	Item 34	0.9%	Item 36	1.0%
	Item 35	1.7%	Item 37	2.3%

6.3 Intercorrelations

The Pearson product-moment correlation between student scores on the Language Arts Literacy and Mathematics content areas for grade 3 was .70; this correlation for grade 4 was also .70. Tables 6.3.1 and 6.3.2 show the correlations between students' scores in the major content clusters and item types. Tables 6.3.3 and 6.3.4 show the correlations between students' scores on the content clusters. The scores used for all correlations were expressed in the raw score metric.

Note that correlations between a content area and cluster within that content area are partially a function of the proportion of the content area that is made up of items from the given cluster. All else being equal, clusters that make up a higher proportion of a content area score will tend to have higher cluster-area correlations. For example, the correlation between Mathematics Total and Mathematics Multiple-Choice in Table 6.3.2 is quite high at .96 because 28 Mathematics Multiple-Choice points are part of the Mathematics Total 43 points.

In addition, correlations are partially a function of the number of items in the measures being correlated; for a given pair of traits, increasing the number of items tends to increase correlations because of the increase in score reliability. Therefore, the number of items in the content areas and clusters being correlated must be considered when their correlations are evaluated.

TABLE 6.3.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Intercorrelations Among Major Content Clusters and Item Types – Grade 3**

Major Content Clusters and Item Types	Major Content Clusters and Item Types							
	Language Arts Literacy (LAL)					Mathematics (MAT)		
	LAL	R	R MC	R OE	W	MAT	M MC	M OE
LAL Language Arts Literacy (40)								
R Reading (20)	.93							
R MC Reading Multiple-Choice (12)	.85	.94						
R OE Reading Open-ended (8)	.80	.82	.58					
W Writing (20)	.85	.61	.52	.59				
MAT Mathematics (33)	.70	.69	.65	.57	.54			
M MC Mathematics Multiple-Choice (24)	.68	.68	.65	.55	.52	.96		
M OE Mathematics Open-ended (9)	.55	.53	.48	.47	.45	.82	.62	

Number in Parentheses is the number of points.
Language Arts Literacy N=103,407; Mathematics N=103,551.

TABLE 6.3.2

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Intercorrelations Among Major Content Clusters and Item Types – Grade 4**

Major Content Clusters and Item Types	Major Content Clusters and Item Types							
	Language Arts Literacy (LAL)					Mathematics (MAT)		
Major Content Clusters and Item Types	LAL	R	R MC	R OE	W	MAT	M MC	M OE
LAL Language Arts Literacy (43)								
R Reading (23)	.94							
R MC Reading Multiple-Choice (11)	.83	.91						
R OE Reading Open-ended (12)	.87	.88	.61					
W Writing (20)	.88	.66	.53	.67				
MAT Mathematics (43)	.70	.70	.64	.61	.55			
M MC Mathematics Multiple-Choice (28)	.67	.67	.62	.58	.53	.96		
M OE Mathematics Open-ended (15)	.63	.64	.58	.56	.50	.92	.77	

Number in Parentheses is the number of points.
Language Arts Literacy N=103,815; Mathematics N=103,767.

TABLE 6.3.3

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Intercorrelations Among Content Areas and Clusters – Grade 3**

Test Section/Cluster	Test Section/Cluster													
	LAL Language Arts Literacy							MAT Mathematics						
Test Section/Cluster	LAL	L1	L2	L3	L4	L5	L6	L7	MAT	M1	M2	M3	M4	M5
LAL Language Arts Literacy (40)														
L1 Reading (20)	.93													
L2 Writing (20)	.85	.61												
L3 Writing / Picture (10)	.78	.58	.88											
L4 Writing / Poem (10)	.73	.50	.89	.57										
L5 Working with Text (9)	.83	.91	.51	.50	.41									
L6 Analyzing Text (11)	.86	.90	.59	.56	.49	.64								
L7 Read First (8)	.82	.89	.51	.50	.41	.97	.64							
MAT Mathematics (33)	.70	.69	.54	.52	.44	.64	.62	.63						
M1 Number Sense and Numerical Operations (9)	.60	.59	.47	.45	.38	.55	.52	.55	.85					
M2 Geometry and Measurement (8)	.53	.53	.41	.40	.33	.48	.48	.48	.79	.56				
M3 Data Analysis, Probability and Discrete Math (8)	.63	.62	.49	.46	.40	.57	.55	.56	.86	.64	.57			
M4 Patterns and Algebra (8)	.56	.55	.43	.41	.35	.50	.49	.50	.82	.61	.52	.61		
M5 Problem Solving (25.5)	.69	.68	.54	.51	.44	.62	.61	.62	.98	.81	.78	.84	.84	

Number in Parentheses is the number of points.
Language Arts Literacy N=103,407; Mathematics N=103,551.

TABLE 6.3.4

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Intercorrelations Among Content Areas and Clusters – Grade 4**

Test Section/Cluster	Test Section/Cluster												
	LAL Language Arts Literacy						MAT Mathematics						
Test Section/Cluster	LAL	L1	L2	L3	L4	L5	L6	MAT	M1	M2	M3	M4	M5
LAL Language Arts Literacy (43)													
L1 Reading (23)	.94												
L2 Writing (20)	.88	.66											
L3 Writing / Picture (10)	.80	.61	.89										
L4 Writing / Poem (10)	.77	.58	.89	.60									
L5 Working with Text (8)	.74	.82	.48	.45	.41								
L6 Analyzing Text (15)	.91	.96	.66	.61	.58	.62							
MAT Mathematics (43)	.70	.70	.55	.51	.47	.59	.66						
M1 Number Sense and Numerical Operations(13)	.63	.62	.50	.47	.43	.52	.59	.88					
M2 Geometry and Measurement (10)	.56	.57	.44	.41	.38	.49	.54	.83	.64				
M3 Data Analysis, Probability and Discrete Math (10)	.59	.58	.47	.44	.41	.48	.56	.85	.65	.63			
M4 Patterns and Algebra (10)	.61	.62	.47	.43	.40	.53	.58	.87	.70	.65	.65		
M5 Problem Solving (33)	.68	.68	.53	.49	.45	.58	.64	.99	.85	.82	.84	.88	

Number in Parentheses is the number of points.

Language Arts Literacy N=103,815; Mathematics N=103,767.

6.4 Item Bias Statistics

Following the classical item analyses, Differential Item Functioning (DIF) studies were completed. One of the goals of test development is to assemble a set of items that provides an estimate of a student's ability that is as fair and accurate as possible for all groups within the population. DIF statistics are used to identify those items that identifiable groups of students (e.g. females, African Americans, Hispanics) with the same underlying level of ability have different probabilities of answering correctly. If the item is differentially more difficult for an identifiable subgroup, the item may be measuring something different from the intended construct. However, it is important to recognize that DIF flagged items might be related to actual differences in relevant knowledge or skill (item impact) or statistical Type I error. As a result, DIF statistics are used to identify potential sources of item bias. Subsequent review by content experts and bias/sensitivity committees determines the source and meaning of any differences that are seen.

ETS used two DIF detection methods: the Mantel-Haenszel and standardization approaches. As part of the Mantel-Haenszel procedure, the statistic described by Holland & Thayer (1986), known as MH D-DIF, was used. This statistic is expressed as the differences between the focal and reference group performance after conditioning on total test score. This statistic is reported on the ETS delta scale, which is a normalized transformation of item difficulty (proportion correct) with a mean of 12 and a standard deviation of 4. Negative MH D-DIF statistics favor the reference group and positive values favor the focal group. The classification logic used for flagging items is based on a combination of absolute differences and significance testing. Items that are not statistically significantly different based on the MH D-DIF ($p > 0.05$) are considered to have similar performance between the two studied groups; these items are considered to be functioning appropriately. For items where the statistical test indicates significant differences ($p < 0.05$), the effect size is used to determine the direction and severity of the DIF. For the LAL OE items, the Mantel-Haenszel procedure was executed where item categories are treated as integer scores and a chi-square test was carried out with one degree of freedom. The male and white groups are considered as reference groups and the female and other ethnic groups are categorized as focal groups.

Based on these DIF statistics, items are classified into one of three categories and assigned values of A, B or C (see Table 6.4.1). Category A contains negligible DIF, Category B items exhibit slight or moderate DIF, and Category C items have moderate to large values of DIF. Negative values imply that conditional on the matching variable, the focal group has a lower mean item score than the reference group. In contrast a positive value implies that, conditional on the matching variable, the reference group has lower mean item score than the focal group. For constructed-response items the MH D-DIF is not calculated, but analogous flagging rules based on the chi-square statistic are applied, resulting in classification into A, B, or C DIF categories.

TABLE 6.4.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
DIF Categories**

DIF Category	Definition
A (negligible)	MH D-DIF not significantly different from zero, or has an absolute value less than one.
B (slight to moderate)	MH D-DIF is significantly different from zero, and is either a) less than 1.5, or b) not significantly different from one.
C (moderate to large)	MH D-DIF is significantly different from one, and has an absolute value greater than 1.5.

Operational items flagged for negative C (C-)DIF are reviewed by an expert DIF review panel consisting of NJDOE staff responsible for the NJ ASK, and external educators identified by NJDOE during the item review meetings, to ensure that the items are free from any bias before being used to produce final test scores.

PART 7: SCALING AND EQUATING

When tests are administered on multiple occasions, there is a need to create multiple forms. A test form is a set of test questions that is built according to a set of content and statistical test specifications (Millman and Greene, 1989). It is difficult to create two forms that are identical in difficulty. Kolen and Brennan (1995) define equating as a statistical process used to adjust scores on test forms so scores on the forms can be used interchangeably. For example, the level of knowledge and skills need to obtain a score of 200 on the 2004 grade 4 NJ ASK Mathematics form must be the same level of knowledge and skills needed to obtain a 200 on the 1999 grade 4 NJ ASK Mathematics form. To facilitate the correct interpretation of scores from multiple forms, test scores are reported as scaled scores. Each form of a test has its own raw-to-scale conversion. The scale scores are intended to be comparable across forms within a grade and subject. NJ ASK scale scores are not comparable across subjects (e.g., LAL and Math) or grades (e.g., 3 and 4).

7.1 Scale Scores

The total scores in the 2004 NJ ASK Language Arts Literacy and Mathematics sections are reported as scale scores with a range of 100 to 300. Please note that 100 and 300 are a theoretical floor and ceiling and may not actually be observed. The scale score of 200 is the cut point between Partially Proficient and Proficient students. The scale score of 250 is the cut point between Proficient and Advanced Proficient students. The score ranges are as follows:

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

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

Scale scores for the NJ ASK tests are linearly related to the raw score metric of the base year. Thus, to obtain scale scores for each test, a set of scaling parameters are applied to the raw score metrics in the base years. The base year is the year the cut scores were set on the form. The base year for the grade 4 Language Arts Literacy test is 2001. For grade 4 Mathematics, the base year is 1999. For grade 3, 2004 is the base year. Table 7.1.1 shows the scaling parameters for each test.

TABLE 7.1.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Scaling Parameters for Base Forms**

Grade	Subject	Base Year	Points	Slope	Intercept
3	Language Arts Literacy	2004	0-40	4.00000	128.0000
	Mathematics	2004	0-33	4.76190	119.0477
4	Language Arts Literacy	2001	0-43	4.34783	106.52174
	Mathematics	1999	0-43	4.16667	104.16659

7.2 Equating Language Arts Literacy

Since 2004 is the base year for grade 3 tests, no equating was needed for the 2004 grade 3 Language Arts Literacy (LAL) test. Scores on the 2004 NJ ASK grade 4 LAL form were equated back to scores on the 2001 LAL base form via 2003 anchored Rasch difficulty parameters and using IRT true score equating procedures. The grade 4 base year LAL raw score scale ranged from 0-43.0. The base year raw cut score for Proficient was 21.5 (200) and the raw cut score for Advanced Proficient was 33.0 (250). These raw cut scores were derived from a standard-setting workshop in 2001.

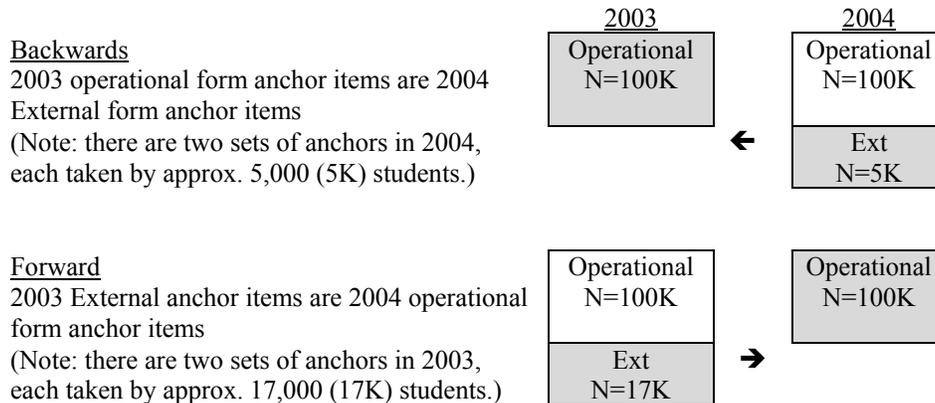
To perform equating, data must be collected. NJ ASK uses a Common-Item Nonequivalent Groups design. Common items are items that appear on both the reference (e.g., 2003) and new (e.g., 2004) forms. Common items are often also called linking and/or anchor items. The meaning of "Nonequivalent groups" is that a different set of students took the reference and new forms, and no assumptions are made that the two groups are equal in ability. The groups could have the same ability, but the students taking the new form could also be more able or less able than the students taking the reference form.

The Language Arts Literacy equating design makes use of external anchor items (i.e., common items that do not count toward a student's operational score). LAL uses an external anchor design that allows for two sets of anchor items to be used in the equating. The two designs have

been called Backwards and Forward. The Backwards equating anchor items were operational items on the old form (e.g., 2003) and are in external sets on the new form (e.g., 2004). The Forward equating items were “pre-tested” as external sets on the old form (2003) and appear in the operational form on the new form (2004).

Figure 7.2.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
LAL Backwards and Forward Equating Designs**



Performance on the Forward equating anchor items in 2004 indicate students in 2004 were more able than in 2003, and the 2004 form was slightly less difficult than the 2003 form. After comparing the results of these two equating approaches, the recommended raw-score to scale-score conversion for the 2004 NJ ASK LAL test resulted from the Forward approach. The recommended raw score cut points in 2004 for LAL were 19.0 and 33.5 for Proficient and Advanced Proficient categories, respectively. Details about the methods and results are described in the 2004 LAL Equating Report. Table 7.2.1 shows the Rasch difficulty parameters (“Measure”), and item fit statistics from WINSTEPS for the Forward equating solution. Table 7.2.2 shows the fixed step parameters for the open-ended anchor items. The raw-to-scale score conversion tables for Language Arts Literacy for 2004 may be found in Appendix C. To create a Braille form a committee reviewed the 2004 Language Arts Literacy test items. Items that could not be translated into Braille were dropped from the Braille version of the operational form. A separate raw-to-scale score conversion table was created for the Braille form.

TABLE 7.2.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Language Arts Literacy Item Parameters – Grade 4**

Item No.	Measure	Anchor	Error	INFIT		OUTFIT		Score	
				MNSQ	ZSTD	MNSQ	ZSTD	Corr.	Displace
1	0.4161	Free	.0016	.89	-9.9	.91	-9.9	.79	-.0005
2	-0.3713	Anchor	.0043	1.11	9.9	1.11	9.9	.43	.1317
3	-0.6137	Anchor	.0049	1.01	2.2	.87	-9.9	.46	.1044
4	-0.0258	Anchor	.0038	.97	-8.9	.97	-4.0	.45	-.0248
5	-0.0055	Anchor	.0038	1.03	7.2	1.01	0.8	.44	.0348
6	-0.8409	Anchor	.0056	.91	-9.9	.69	-9.9	.41	-.0015
7	0.7153	Anchor	.0024	.90	-9.9	.91	-9.9	.73	-.1939
8	0.8845	Anchor	.0022	.76	-9.9	.75	-9.9	.72	-.0478
9	0.4197	Free	.0016	.95	-9.9	.96	-7.9	.77	-.0007
10	0.1679	Anchor	.0037	1.15	9.9	1.24	9.9	.36	.0842
11	-0.2232	Anchor	.0041	1.07	9.9	1.16	9.9	.34	-.0246
12	-0.5632	Anchor	.0047	1.01	2.5	1.10	7.9	.33	-.0244
13	0.2845	Anchor	.0036	.96	-9.9	.98	-4.5	.48	-.0097
14	0.3948	Anchor	.0036	1.16	9.9	1.27	9.9	.32	-.0203
15	0.2490	Anchor	.0036	1.20	9.9	1.33	9.9	.36	.2438
16	0.5375	Anchor	.0022	.60	-9.9	.61	-9.9	.66	.0679

TABLE 7.2.2

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Forward Language Arts Literacy Fixed OE Item Step Parameters – Grade 4**

Item	Category	Step	Item	Category	Step	Item	Category	Step
7	0	.00	8	0	.00	16	0	.00
7	1	-1.06	8	1	-.66	16	1	-.50
7	2	-2.03	8	2	-1.65	16	2	-1.58
7	3	-.58	8	3	-.61	16	3	-.69
7	4	-.49	8	4	-.76	16	4	-.56
7	5	.46	8	5	.40	16	5	.22
7	6	.62	8	6	.47	16	6	.59
7	7	1.35	8	7	1.22	16	7	1.18
7	8	1.73	8	8	1.58	16	8	1.33

7.3 Equating Mathematics

Since 2004 is the base year for grade 3 tests, no equating was needed for the 2004 grade 3 Mathematics test. Scores on the 2004 NJ ASK grade 4 Mathematics form were equated back to scores on the 1999 Mathematics base form via 2003 anchored Rasch difficulty parameters and using IRT true score equating procedures. The grade 4 base year Mathematics raw score scale ranged from 0-43.0. The base year raw cut score for Proficient was 23.0 (200) and the raw cut score for Advanced Proficient was 35.0 (250). These raw cut scores were derived from a standard-setting workshop in 2001.

The data collection design for the NJ ASK Mathematics test is also Common-Item Nonequivalent Groups design. The Mathematics test uses internal anchor items. Internal anchor items are common items that are embedded in the operational set of items (i.e., they count toward a student's operational score). In 2004, 11 items from the 2003 operational form also appeared in the 2004 operational form. Ten anchor items were multiple-choice and 1 was open-ended.

Based on the performance on the anchor items, the 2004 students appear to be more able than the 2003 students and the 2004 form was more difficult than previous math forms. The recommended raw-score (and scale-score) cut points for the 2004 Mathematics NJ ASK based on the equating results were 19.5 (200) and 32.5 (250) for Proficient and Advanced Proficient categories, respectively. Details about the methods and results are described in the 2004 NJ ASK Mathematics Equating Report. Table 7.3.1 shows the Rasch difficulty parameters and item fit statistics from WINSTEPS for the equating. Table 7.3.2 shows the fixed step parameters for the open-ended items. To create a Braille form a committee reviewed the 2004 Mathematics test items. Items that could not be translated into Braille were dropped from the Braille version of the operational form. A separate raw-to-scale score conversion table was created for the Braille form.

TABLE 7.3.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Mathematics Item Parameters – Grade 4**

Item No.	Measure	Anchor	Error	IN FIT		OUT FIT		Score Corr.	Displace
				MNSQ	ZSTD	MNSQ	ZSTD		
1	-2.5260	Free	0.0169	0.95	-3.3	0.70	-9.9	0.24	-0.0038
2	-1.0050	Free	0.0092	0.91	-9.9	0.80	-9.9	0.36	-0.0046
3	-0.4966	Free	0.0080	0.89	-9.9	0.82	-9.9	0.42	-0.0048
4	-0.2171	Free	0.0075	0.87	-9.9	0.81	-9.9	0.46	-0.0047
5	-1.5845	Anchor	0.0114	0.89	-9.9	0.71	-9.9	0.32	-0.0491
6	-0.9482	Anchor	0.0091	0.95	-8.3	0.89	-9.9	0.29	-0.0009
7	-0.3531	Free	0.0077	0.92	-9.9	0.87	-9.9	0.38	-0.0046
8	0.0790	Free	0.0071	0.87	-9.9	0.84	-9.9	0.46	-0.0052
9	0.2233	Anchor	0.0040	1.24	9.9	1.38	9.9	0.28	0.0972
10	0.3831	Anchor	0.0038	1.13	9.9	1.19	9.9	0.35	0.0590
11	0.1944	Anchor	0.0041	1.14	9.9	1.23	9.9	0.36	0.1074
12	0.0178	Free	0.0043	1.08	9.9	1.19	9.9	0.31	-0.0046
13	0.2823	Free	0.0039	1.04	9.9	1.06	8.6	0.38	-0.0049
14	-0.0751	Free	0.0045	0.98	-4.3	0.95	-4.5	0.38	-0.0047
15	0.8113	Anchor	0.0036	0.88	-9.9	0.84	-9.9	0.54	0.1305
16	0.6612	Anchor	0.0037	1.00	1.2	1.01	2.1	0.46	0.1275
17	1.1627	Free	0.0038	0.96	-9.9	1.03	5.5	0.46	-0.0050
18	-0.2557	Free	0.0050	1.00	0.2	0.96	-3.4	0.34	-0.0046
19	0.9414	Free	0.0037	0.97	-9.9	0.97	-5.6	0.46	-0.0050
20	0.6414	Free	0.0037	1.06	9.9	1.10	9.9	0.38	-0.0048
21	0.7859	Free	0.0019	1.14	9.9	1.17	9.9	0.62	-0.0049
22	0.3317	Anchor	0.0039	0.95	-9.9	0.94	-9.5	0.37	-0.1844
23	-0.0892	Free	0.0046	0.93	-9.9	0.83	-9.9	0.42	-0.0046
24	0.6845	Free	0.0037	1.06	9.9	1.07	9.9	0.39	-0.0049
25	0.9746	Free	0.0037	1.10	9.9	1.16	9.9	0.36	-0.0050
26	0.3802	Anchor	0.0038	0.98	-7.2	0.94	-9.3	0.47	0.0681
27	0.7619	Anchor	0.0037	0.92	-9.9	0.91	-9.9	0.50	-0.0065
28	0.8805	Free	0.0018	0.86	-9.9	0.82	-9.9	0.71	-0.0051
29	0.9708	Free	0.0018	1.18	9.9	1.24	9.9	0.63	-0.0056
30	1.0033	Free	0.0037	1.06	9.9	1.17	9.9	0.38	-0.0050
31	0.1669	Free	0.0041	1.01	2.5	1.04	5.3	0.38	-0.0044
32	0.6150	Free	0.0037	1.07	9.9	1.09	9.9	0.38	-0.0047
33	0.6876	Free	0.0037	1.07	9.9	1.10	9.9	0.38	-0.0048
34	0.4580	Free	0.0038	0.97	-9.6	0.96	-6.9	0.45	-0.0047
35	0.1068	Free	0.0042	0.95	-9.9	0.91	-9.9	0.43	-0.0049
36	0.3786	Anchor	0.0021	1.00	0.7	1.11	9.9	0.61	-0.0230
37	1.3034	Free	0.0021	1.07	9.9	1.05	6.5	0.62	-0.0057

TABLE 7.3.2

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Mathematics Fixed OE Item Step Parameters – Grade 4**

Item	Category	Step	Item	Category	Step	Item	Category	Step
21	0	0.00	28	0	0.00	29	0	0.00
21	1	1.47	28	1	1.66	29	1	1.27
21	2	-1.49	28	2	-0.97	29	2	-0.98
21	3	1.39	28	3	0.71	29	3	1.33
21	4	-1.86	28	4	-1.28	29	4	-0.46
21	5	2.67	28	5	1.81	29	5	0.52
21	6	-2.18	28	6	-1.93	29	6	-1.69
Item	Category	Step	Item	Category	Step			
36	0	0.00	37	0	0.00			
36	1	1.41	37	1	1.78			
36	2	-2.01	37	2	-2.05			
36	3	1.24	37	3	0.82			
36	4	-0.70	37	4	-0.35			
36	5	1.27	37	5	0.20			
36	6	-1.21	37	6	-0.40			

PART 8: TEST STATISTICS

8.1 Summary Statistics

Means and standard deviations of students' raw scores on each content area are given in Tables 8.1.1 (grade 3) and 8.1.2 (grade 4) for the March 2004 test. These data are based on the total student populations described in Part 1 and Appendix A. Table 8.1.1 shows that grade 3 students' mean raw scores were 21.9 of 40 points for Language Arts Literacy, and 21.7 of 33 points for Mathematics. The table also shows the standard deviations of the raw scores for grade 3 were 5.7 on Language Arts Literacy and 6.6 on Mathematics. Table 8.1.2 shows that grade 4 students' mean raw scores were 24.2 of 43 points for Language Arts Literacy, and 25.2 of 43 points for Mathematics. The table also shows the standard deviations of the raw scores for grade 4 were 6.2 on Language Arts Literacy and 9.0 on Mathematics. Raw score to scale score conversion tables by content area are included in Appendix C. Also, frequency distributions of the scale scores by content area are shown in Appendix C.

TABLE 8.1.1**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Means and Standard Deviations of Students'
Raw Scores by Test Section – Grade 3**

TEST SECTION	Number of Points	Raw Scores Mean	Standard Deviation	Number Tested
Language Arts Literacy	40	21.9	5.7	103,414
Mathematics	33	21.7	6.6	103,559

TABLE 8.1.2**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Means and Standard Deviations of Students'
Raw Scores by Test Section – Grade 4**

TEST SECTION	Number of Points	Raw Scores Mean	Standard Deviation	Number Tested
Language Arts Literacy	43	24.2	6.2	103,818
Mathematics	43	25.2	9.0	103,770

Means and Standard Deviations of Students' Raw Scores

Tables 8.1.3 and 8.1.4 report the means and standard deviations for students' obtained numbers of raw score points by cluster on the March 2004 tests. Table 8.1.3 shows that in Language Arts Literacy, grade 3 students' mean percent correct was 54.7% overall with 61.2% in Reading and 48.2% in Writing. The mean raw score on the writing/speculate task in response to a picture was 5.0 points out of a possible 10 points and the mean raw score on the writing/analyze task in response to a poem was 4.7 points out of a possible 10 points. The mean percents correct in the two Reading clusters—Working with Text and Analyzing/Critiquing Text—was 73.1% and 51.5%.

With respect to the grade 3 students' percent correct scores on the Mathematics content clusters, the data in Table 8.1.3 indicate that the mean percent correct ranged from 61.1% in Geometry and Measurement to 72.3% in Number Sense and Numerical Operations. The mathematics items are also categorized as Problem Solving and Total. The mean percent correct was 63.1% for Problem Solving and 65.6% for Total.

Table 8.1.4 shows that in Language Arts Literacy, grade 4 students' mean percent correct was 56.2% overall with 57.8% in Reading and 54.4% in Writing. The mean raw score on the writing/speculate task in response to a picture was 5.6 points out of a possible 10 points and the mean raw score on the writing/analyze task in response to a poem was 5.3 points out of a possible 10 points. The mean percents correct in the two Reading clusters—Working with Text and Analyzing/Critiquing Text—were 72.5% and 52.7%.

With respect to the grade 4 students' percent correct scores on the Mathematics content clusters, the data in Table 8.1.4 indicate that the mean percent correct ranged from 54.8% in Patterns and Algebra to 66.3% in Number Sense and Numerical Operations. The mathematics items are also categorized as Problem Solving and Total. The mean percent correct was 54.3% for Problem Solving and 58.7% for Total. Tables 8.1.5 and 8.1.6 show the means and standard deviations for the students' raw scores and percent correct scores on the dichotomously scored items by NJ ASK Content Area. Tables 8.1.7 and 8.1.8 provide means and standard deviations for students' raw scores and percent correct scores on the open-ended items by cluster.

TABLE 8.1.3

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Means and Standard Deviations of Students' Raw Scores
and Percent Correct by Content Area – Grade 3**

NJ ASK Content Area	Number of Items		Number of Possible Points	Raw Score		Percent Correct	
	Multiple-Choice	Open-Ended		Raw Scores Mean	Standard Deviation	Mean	Standard Deviation
Language Arts Literacy	12	4	40	21.9	5.7	54.7	14.3
Writing	0	2	20	9.6	2.6	48.2	13.1
Writing/Picture	0	1	10	5.0	1.5	49.7	14.5
Writing/Poem	0	1	10	4.7	1.5	46.6	15.0
Reading	12	2	20	12.2	3.7	61.2	18.7
Working with Text	9	0	9	6.6	2.1	73.1	23.3
Analyzing Text	3	2	11	5.7	2.0	51.5	18.6
Mathematics*	27	3	33	21.7	6.6	65.6	19.9
Number Sense and Numerical Operations*	12	0	9	6.5	2.0	72.3	22.6
Geometry and Measurement	5	1	8	4.9	1.8	61.1	22.8
Patterns and Algebra	5	1	8	5.2	1.9	65.3	24.3
Data Analysis, Probability, and Discrete Math	5	1	8	5.0	2.1	63.1	26.4
Problem Solving	17	3	25.5	16.1	5.4	63.1	21.0

* Six multiple-choice items in the Number Sense and Numerical Operations cluster and in the Mathematics total raw score are counted as one-half point.

TABLE 8.1.4

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Means and Standard Deviations of Students' Raw Scores
and Percent Correct by Content Area – Grade 4**

NJ ASK Content Area	Number of Items		Number of Possible Points	Raw Score		Percent Correct	
	Multiple- Choice	Open- Ended		Raw Scores Mean	Standard Deviation	Mean	Standard Deviation
Language Arts Literacy	11	5	43	24.2	6.2	56.2	14.4
Writing	0	2	20	10.9	2.8	54.4	14.1
Writing/Picture	0	1	10	5.6	1.6	55.6	15.8
Writing/Poem	0	1	10	5.3	1.6	53.2	15.7
Reading	11	3	23	13.3	4.0	57.8	17.2
Working with Text	6	0	6	4.4	1.4	72.5	24.1
Analyzing Text	5	3	17	9.0	2.9	52.7	17.1
Mathematics*	32	5	43	25.2	9.0	58.7	21.0
Number Sense and Numerical Operations*	11	2	13	8.6	2.9	66.3	22.6
Geometry and Measurement	7	1	10	5.6	2.3	56.3	22.6
Patterns and Algebra	7	1	10	5.5	2.7	54.8	26.9
Data Analysis, Probability, and Discrete Math	7	1	10	5.5	2.6	55.0	25.8
Problem Solving	18	5	33	17.9	7.5	54.3	22.6

* Eight multiple-choice items in the Number Sense and Numerical Operations cluster and in the Mathematics total raw score are counted as one-half point.

TABLE 8.1.5

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Means and Standard Deviations of Students' Raw Scores
and Percent Correct on the Dichotomously Scored Items
by Content Area – Grade 3**

NJ ASK Content Area	Number of Points	Raw Scores		Percent Correct	
		Mean	Standard Deviation	Mean	Standard Deviation
Language Arts Literacy	12	8.5	2.6	71.1	22.1
Writing ^a	--	--	--	--	--
Writing/Picture	--	--	--	--	--
Writing/Poem	--	--	--	--	--
-----	-----	-----	-----	-----	-----
Reading	12	8.5	2.6	71.1	22.1
Working with Text	9	6.6	2.1	73.1	23.3
Analyzing Text	3	2.0	0.9	65.1	29.3
Mathematics*	24	17.0	4.9	70.8	20.3
Number Sense and Numerical Operations*	9	6.5	2.0	72.3	22.6
Geometry and Measurement	5	3.9	1.1	77.2	22.4
Patterns and Algebra	5	3.3	1.5	65.9	29.1
Data Analysis, Probability and Discrete Math	5	3.3	1.4	66.8	28.0
-----	-----	-----	-----	-----	-----
Problem Solving	16.5	11.4	3.6	69.2	22.0

* Six items in the Number Sense and Numerical Operations cluster and in the Mathematics total raw score are counted as one-half point.

a. There were no dichotomously scored writing items.

TABLE 8.1.6

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Means and Standard Deviations of Students' Raw Scores
and Percent Correct on the Dichotomously Scored Items
by Content Area –Grade 4**

NJ ASK Content Area	Number of Points	Raw Scores		Percent Correct	
		Mean	Standard Deviation	Mean	Standard Deviation
Language Arts Literacy	11	7.6	2.4	69.5	21.6
Writing ^a	--	--	--	--	--
Writing/Picture	--	--	--	--	--
Writing/Poem	--	--	--	--	--
-----	-----	-----	-----	-----	-----
Reading	11	7.6	2.4	69.5	21.6
Working with Text	6	4.4	1.4	72.5	24.1
Analyzing Text	5	3.3	1.3	65.8	25.6
Mathematics*	7	5.0	1.5	71.5	21.9
Number Sense and Numerical Operations*	7	4.8	1.6	68.0	23.0
Geometry and Measurement	7	4.1	1.8	58.2	25.6
Patterns and Algebra	7	4.3	1.8	61.1	25.2
Data Analysis, Probability and Discrete Math	28	18.1	5.5	64.7	19.5
-----	-----	-----	-----	-----	-----
Problem Solving	11	7.6	2.4	69.5	21.6

* Eight items in the Number Sense and Numerical Operations cluster and in the Mathematics total raw score are counted as one-half point.

a. There were no dichotomously scored writing items.

TABLE 8.1.7**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Means and Standard Deviations of Students' Raw Scores
and Percent Correct on the Open-Ended Items by Content Areas and Clusters – Grade 3**

NJ ASK Content Area	Number		Raw Scores		Percent Correct	
	Items	Points	Mean	Standard Deviation	Mean	Standard Deviation
Language Arts Literacy	4	28	13.3	3.7	47.7	13.3
Writing	2	20	9.6	2.6	48.2	13.1
Writing/Picture	1	10	5.0	1.5	49.7	14.5
Writing/Poem	1	10	4.7	1.5	46.6	15.0
Reading	2	8	3.7	1.5	46.4	19.1
Working with Text	0	0	--	--	--	--
Analyzing Text	2	8	3.7	1.5	46.4	19.1
Mathematics	3	9	4.7	2.3	51.8	25.8
Number Sense, and Numerical Operations	0	0	--	--	--	--
Geometry and Measurement	1	3	1.0	1.1	34.2	37.6
Patterns and Algebra	1	3	1.9	0.9	64.3	31.1
Data Analysis Probability and Discrete Math	1	3	1.7	1.1	56.9	36.2
Problem Solving	3	9	4.7	2.3	51.8	25.8

TABLE 8.1.8**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Means and Standard Deviations of Students' Raw Scores
and Percent Correct on the Open-Ended Items by Content Areas and Clusters – Grade 4**

NJ ASK Content Area	Number		Raw Scores		Percent Correct	
	Items	Points	Mean	Standard Deviation	Mean	Standard Deviation
Language Arts Literacy	5	32	16.5	4.4	51.7	13.9
Writing	2	20	10.9	2.8	54.4	14.1
Writing/Picture	1	10	5.6	1.6	55.6	15.8
Writing/Poem	1	10	5.3	1.6	53.2	15.7
Reading	3	12	5.7	2.0	47.2	17.0
Working with Text	0	0	--	--	--	--
Analyzing Text	3	12	5.7	2.0	47.2	17.0
Mathematics	5	15	7.1	4.1	47.4	27.4
Number Sense, and Numerical Operations	2	6	3.6	1.8	60.2	30.4
Geometry and Measurement	1	3	0.9	1.0	29.1	34.7
Patterns and Algebra	1	3	1.4	1.3	46.8	41.9
Data Analysis Probability and Discrete Math	1	3	1.2	1.3	40.7	41.9
Problem Solving	5	15	7.1	4.1	47.4	27.4

8.2 Classical Reliability Estimates of the Test Scores

Tables 8.2.1 and 8.2.2 summarize reliability estimates for the NJ ASK grades 3 and 4 content areas and clusters. The reliability coefficients given in these tables are based on Cronbach's coefficient alpha measure of internal consistency. Cronbach's alpha is used on tests containing items that can be scored along a range of values. The standard errors of measurement (SEMs) for the major content areas - Language Arts Literacy and Mathematics - are expressed in terms of the raw score metric and the scale score metric. The NJ ASK scale scores range from 100 to 300.

Reliabilities and SEMs for the dichotomously scored items in each cluster are reported in Tables 8.2.3 and 8.2.4.

When evaluating these results, it is important to recall that reliability is partially a function of test length. Therefore, the reliability of a content area is likely to be greater than the reliability of a

cluster simply because the content area has more items. Similarly, clusters with more items are likely to be more reliable than clusters with fewer items. The data provided in Tables 8.2.1, 8.2.2, 8.2.3 and 8.2.4 reflect the expected positive relationship between test length and reliability.

The SEMs given in Tables 8.2.1, 8.2.2, 8.2.3 and 8.2.4 are useful when interpreting students' scores. Measurement error occurs in every test. A student's true score is a hypothetical average score that the student would obtain if a test were repeatedly administered to the student without the effects of instruction, practice, or fatigue. Mehrens and Lehmann (1991) suggest this use of the SEM:

The standard error of measurement is often used for what is called band interpretation. Band interpretation helps convey the idea of imprecision of measurement.... If we assume that the errors are random, an individual's observed scores will be normally distributed about his true score over repeated testing. Thus, one can say that a person's observed score will lie between ± 1 SE of his true score approximately 68 percent of the time, or ± 2 SE of his true score about 95 percent of the time (p. 252).

TABLE 8.2.1

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

**Reliability Estimates and Standard Errors of Measurement (SEM)
for Content Areas and Clusters – Grade 3**

NJ ASK Test Section	Number of Points	Reliability	Raw Score SEM	Scale Score SEM
Language Arts Literacy	40	0.84	2.28	9.12
Reading	20	0.80	1.66	
Writing	20	0.72	1.37	
Working with Text	9	0.70	1.14	
Analyzing Text	11	0.65	1.21	
Mathematics	33	0.86	2.45	11.65
Number Sense and Numerical Operations	9	0.71	1.10	
Geometry and Measurement	8	0.44	1.37	
Patterns and Algebra	8	0.57	1.28	
Data analysis, Probability and Discrete Math	8	0.60	1.33	
Problem Solving	25.5	0.83	2.23	

TABLE 8.2.2**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Reliability Estimates and Standard Errors of Measurement (SEM)
for Content Areas and Clusters – Grade 4**

NJ ASK Test Section	Number of Points	Reliability	Raw Score SEM	Scale Score SEM
Language Arts Literacy	43	0.85	2.38	8.97
Reading	23	0.82	1.68	
Writing	20	0.75	1.41	
----- Working with Text	6	0.55	0.97	
Analyzing Text	17	0.78	1.38	
Mathematics	43	0.89	2.93	11.21
Number Sense and Numerical Operations	13	0.73	1.47	
Geometry and Measurement	10	0.59	1.44	
Patterns and Algebra	10	0.60	1.71	
Data analysis, Probability and Discrete Math	10	0.61	1.62	
----- Problem Solving	33	0.86	2.77	

TABLE 8.2.3**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Reliability Estimates and Standard Errors of Measurement (SEM)
for Dichotomously Scored Items Within Content Clusters – Grade 3**

NJ ASK Content Area	Number of Points	Reliability	Raw Score SEM
Language Arts Literacy	12	0.74	1.34
Reading	12	0.74	1.34
Writing*	--	--	--
Writing/Picture	--	--	--
Writing/Poem	--	--	--
-----	-----	-----	-----
Working with Text	9	0.70	1.14
Analyzing Text	3	0.36	0.70
Read First	8	0.71	1.04
Mathematics	24	0.84	1.92
Number Sense and Numerical Operations	9	0.71	1.10
Geometry and Measurement	5	0.41	0.86
Patterns and Algebra	5	0.59	0.93
Data analysis, Probability and Discrete Math	5	0.59	0.90
-----	-----	-----	-----
Problem Solving	16.5	0.80	1.63

TABLE 8.2.4**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)****Reliability Estimates and Standard Errors of Measurement (SEM)
for Dichotomously Scored Items Within Content Clusters – Grade 4**

NJ ASK Content Area	Number of Points	Reliability	Raw Score SEM
Language Arts Literacy	11	0.68	1.34
Reading	11	0.68	1.34
Writing*	--	--	--
Writing/Picture	--	--	--
Writing/Poem	--	--	--
-----	-----	-----	-----
Working with Text	6	0.55	0.97
Analyzing Text	5	0.47	0.93
Mathematics	28	0.85	2.12
Number Sense and Numerical Operations	7	0.65	0.81
Geometry and Measurement	7	0.55	1.08
Patterns and Algebra	7	0.52	1.24
Data analysis, Probability and Discrete Math	7	0.59	1.13
-----	-----	-----	-----
Problem Solving	18	0.76	1.89

* There were no dichotomously scored writing items.

8.3 Reliability of Performance Classifications

Decision accuracy provides an estimate of how reliably a test form classifies students into performance categories. Decision accuracy is estimated by comparing the observed score distribution for a form to a hypothetical true score distribution. The observed score distribution (also called single-form score distribution) is the actual distribution of scores for all test takers on a test form. The true score distribution is hypothetical because true scores cannot be known, although, they can be estimated. A true score is the average of the observed scores for a student obtained over an infinite number of repeated administrations of the same form.

The methodology used for estimating the reliability of classification and decision accuracy is described in Livingston and Lewis (1995) and is implemented using the ETS-proprietary computer program RELCLASS-COMP (Version 4.12). RELCLASS-COMP generates a contingency table that shows the proportion of exact agreement between the two distributions. In Tables 8.3.1 and 8.3.2, the cells showing exact agreement are shaded. The sum of the shaded, diagonal cells represents the estimated proportion correctly classified.

Table 8.3.1: For grade 3 Language Arts Literacy, the estimated proportion correctly classified overall was 0.81. When the decisions were collapsed to below proficient versus proficient and above, the estimated proportion correctly classified was 0.93. For Mathematics, the estimated proportion correctly classified overall was 0.81. When the decisions were collapsed to below proficient versus proficient and above, the estimated proportion correctly classified was 0.91.

Table 8.3.2: For grade 4 Language Arts Literacy, the estimated proportion correctly classified overall was 0.88. When the decisions were collapsed to below proficient versus proficient and above, the estimated proportion correctly classified was 0.94. For Mathematics, the estimated proportion correctly classified overall was 0.83. When the decisions were collapsed to below proficient versus proficient and above, the estimated proportion correctly classified was 0.91.

TABLE 8.3.1

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

Reliability of Classification and Decision Accuracy – Grade 3

Decision Accuracy: Language Arts Literacy

		<i>Observed Score</i>			Observed Total
		Advanced Proficient (30.5-40.0)	Proficient (18.0-30.0)	Partially Proficient (0-17.5)	
<i>True Score</i>	Placement Score				
	Advanced Proficient (30.5-40.0)	0.01	0.09	0.00	0.11
	Proficient (18.0-30.0)	0.02	0.63	0.04	0.68
	Partially Proficient (0-17.5)	0.00	0.03	0.17	0.21
Expected Total		0.03	0.76	0.21	

Estimated Proportion Correctly Classified: Total = 0.81, Proficient & Above = 0.93

Decision Accuracy: Mathematics

		<i>Observed Score</i>			Observed Total
		Advanced Proficient (27.5-33.0)	Proficient (17.0-27.0)	Partially Proficient (0-16.5)	
<i>True Score</i>	Placement Score				
	Advanced Proficient (27.5-33.0)	0.17	0.06	0.00	0.23
	Proficient (17.0-27.0)	0.04	0.45	0.04	0.54
	Partially Proficient (0-16.5)	0.00	0.05	0.19	0.23
Expected Total		0.21	0.56	0.23	

Estimated Proportion Correctly Classified: Total = 0.81, Proficient & Above = 0.91

TABLE 8.3.2

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)

Reliability of Classification and Decision Accuracy – Grade 4

Decision Accuracy Language Arts Literacy

		<i>Observed Score</i>				
		Placement Score	Advanced Proficient (33.5-43)	Proficient (19-33)	Partially Proficient (0-18.5)	Observed Total
<i>True Score</i>	Advanced Proficient (33.5-43)	0.00	0.06	0.00	0.06	
	Proficient (19-33)	0.00	0.73	0.03	0.76	
	Partially Proficient (0-18.5)	0.00	0.03	0.15	0.18	
	Expected Total	0.00	0.82	0.18		

Estimated Proportion Correctly Classified: Total = 0.88, Proficient & Above = 0.94

Decision Accuracy Mathematics Grade 4

		<i>Observed Score</i>				
		Placement Score	Advanced Proficient (32.5-43)	Proficient (19.5-32)	Partially Proficient (0-19.0)	Observed Total
<i>True Score</i>	Advanced Proficient (32.5-43)	0.21	0.05	0.00	0.26	
	Proficient (19.5-32)	0.04	0.39	0.04	0.47	
	Partially Proficient (0-19.0)	0.00	0.05	0.23	0.28	
	Expected Total	0.25	0.49	0.27		

Estimated Proportion Correctly Classified: Total = 0.83, Proficient & Above = 0.91

8.4 Conditional Estimate of Error at Each Cut-Score

When reviewing a cut score, it is important to keep in mind that there is measurement error surrounding that cut score. Measurement error occurs because no instrument measures a student’s level of knowledge and skills precisely. Think of the student who knows the correct answer to an item, but makes a careless arithmetic error or accidentally marks the wrong response. Or think of a student who really does not know the correct answer but who fills in the correct answer purely by chance. These situations require us to calculate a standard error of measurement for each score. For example, let’s say a student scores a 28 (out of 43) and the

standard error of measurement for the score is about 2.0 raw score points. We can be 95% confident that the student’s ability put him in the range of scoring a 28 plus or minus two standard errors of measurement: that is between 24–32.

The WINSTEPS program calculates the standard error of the measure (SEM) at each score point. Unlike the classical standard error of measurement, the value of the SEM using Item Response Theory varies with ability level. The equation for standard error of estimation is given by

$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}} \quad [8.4.1]$$

where $I(\theta)$ is the information function for a test at θ . For the Rasch model using unweighted raw scores, the information provided by a test at θ is the sum of the item information functions at θ (Hambleton, Swaminathan, and Rogers, 1991). Table 8.4.1 shows conditional estimates of error at each cut score for grades 3 and 4 for each subject.

TABLE 8.4.1
2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Conditional Estimate of Error at Each Cut-Score

Grade	Subject	Proficiency Level	Raw Score Cut	Theta Cut	Theta SE	Approximate SE in Raw Points
3	LAL	Proficient	18.0	-0.0476	0.2200	2.5
		Advanced Proficient	30.5	1.5372	0.2870	2.0
	Math	Proficient	17.0	0.2564	0.1900	2.5
		Advanced Proficient	27.5	1.1247	0.2390	2.0
Grade	Subject	Proficiency Level	Raw Score Cut	Theta Cut	Theta SE	Approximate SE in Raw Points
4	LAL	Proficient	19.0	-0.0362	0.2055	2.5
		Advanced Proficient	33.5	1.4503	0.2466	2.0
	Math	Proficient	19.5	0.4941	0.1645	3.0
		Advanced Proficient	32.5	1.1710	0.1734	3.0

8.5 Rater Reliability

Tables 8.5.1 and 8.5.2 show the percentages of writing tasks and open-ended items scored with exact agreement, adjacent agreement, and resolution needed.

The Writing cluster within Language Arts Literacy consists of two writing activities: a writing/speculate task in response to a picture and a writing/analyze task related to a poem. For

these writing tasks, the rubrics used by the raters had score points that ranged from 0 to 5. If two raters assigned scores to a student’s writing task that were not exactly the same or adjacent, a third “expert” rater also read and assigned a score to the student’s response. Of more than 200,000 task responses in grade 3 in March 2004, 63.9% received exactly the same scores by the raters and 33.8% received scores that were adjacent. Thus, a total of 97.7% of the task responses required only two raters. The remaining 2.3% received scores on the Writing Tasks that differed by more than one point and therefore required a third rater (see Table 8.5.1). For grade 4 Language Arts Literacy in March 2004, 55.2% received exactly the same scores by the raters and 40.7% received scores that were adjacent. Thus, a total of 95.8% of the task responses required only two raters. The remaining 4.2% received scores on the Writing Tasks that differed by more than one point and therefore required a third rater (see Table 8.5.1).

The Reading cluster and the Mathematics content areas include open-ended items. For the Reading open-ended items, the rubric used by the raters had score points that ranged from 0 to 4. For the Mathematics items, the rubric ranged from 0 to 3 points. Table 8.5.1 shows that for grade 3 Reading open-ended items, exact agreement was obtained 65.1% of the time. Resolution by a third rater was needed for 1.8% of the responses. For grade 3 Mathematics, exact agreement was obtained 82% of the time and resolution was needed for 1.4% of the task responses. Table 8.5.2 shows that for grade 4 Reading open-ended items, exact agreement was obtained 55.2% of the time. Resolution by a third rater was needed for 4.1% of the responses. For grade 4 Mathematics, exact agreement was obtained 84.6% of the time and resolution was needed for 1.7% of the task responses.

TABLE 8.5.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Consistency Between Raters Scoring Writing Tasks and Open-Ended Items - Grade 3**

Writing Tasks and Open-Ended Items	Percent Raters In Exact Agreement	Percent Raters In Adjacent Agreement	Percent Resolution Needed
Language Arts Literacy	63.9	33.8	2.3
Writing Total	62.8	34.5	2.8
Writing/Picture	67.0	32.8	0.2
Writing/Poem	58.5	36.1	5.3
Reading Total	65.1	33.1	1.8
Open-Ended Item 1	66.9	32.2	0.9
Open-Ended Item 2	63.2	34.0	2.7
Mathematics	82.0	16.6	1.4
Open-Ended Item 1	84.2	14.8	1.0
Open-Ended Item 2	84.7	13.6	1.7
Open-Ended Item 3	77.0	21.3	1.6

TABLE 8.5.2

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Consistency Between Raters Scoring Writing Tasks and Open-Ended Items – Grade 4**

Writing Tasks and Open-Ended Items	Percent Raters In Exact Agreement	Percent Raters In Adjacent Agreement	Percent Resolution Needed
Language Arts Literacy	55.2	40.7	4.2
Writing Total	55.0	40.7	4.3
Writing/Picture	58.3	38.7	3.0
Writing/Poem	51.7	42.6	5.6
Reading Total	55.2	40.6	4.1
Open-Ended Item 1	54.5	40.6	4.9
Open-Ended Item 2	55.7	40.9	3.5
Open-Ended Item 3	55.5	40.4	4.0
Mathematics	84.6	13.7	1.7
Open-Ended Item 1	88.2	10.1	1.7
Open-Ended Item 2	87.4	11.8	0.8
Open-Ended Item 3	83.5	14.5	2.0
Open-Ended Item 4	82.2	16.4	1.5
Open-Ended Item 5	81.8	15.9	2.3

Part 9: Validity

Content and Curricular Validity

The New Jersey Department of Education is developing a comprehensive set of assessments that measure student achievement of the Core Curriculum Content Standards. The validity of the NJ ASK scores is based on the alignment of the NJ ASK assessments to the Core Curriculum Content Standards and the knowledge and skills expected of third- and fourth-grade students.

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

- Evidence based on test content
- Evidence based on internal structure of the test
- Evidence based on relations to other variables
- Evidence based on consequences of testing

For an assessment like NJ ASK, one intended to measure students' performance in relation to the Core Curriculum Content Standards, content validity evidence is primary. Content validity is the most relevant and important source of evidence. The section of this technical report on "Test Development," presents validity evidence based on test content. A description of the test specification development is followed by the procedures for test item development. Details about item writing as well as task, prompt, and passage selection are included. The last section delineates the review work of the New Jersey Assessment Content Committees. Additionally, an external committee is assisting the New Jersey Department of Education by reviewing the assessments to determine how well they measure the knowledge and skills stated in the standards, and by comparing the New Jersey standards with those in other states and countries.

Part 10: Reporting

Scores are reported in two cycles. Cycle I data is considered preliminary. Schools and districts are encouraged to review student information to make sure it is correct and accurate. Schools have the opportunity to make corrections to student information before Cycle II reports are published. For more information about score reports, please see the NJASK Cycle I Score Interpretation Manual and/or the NJASK Cycle II Score Interpretation Manual.

10.1 Cycle I Reports

The Cycle I reports include the following: Student Sticker, Individual Student Report, All Sections Roster, Student Roster, Summary of School Performance, Summary of District Performance, Summary of School Cluster Performance, and Summary of District Cluster Performance. Each Cycle I report is briefly described below.

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 scores in Language Arts Literacy and Mathematics are provided. Designations of the proficiency levels are printed next to the Language Arts Literacy and Mathematics scale scores. Voids, where applicable, are noted.

Individual Student Report

The Individual Student Report (ISR) is a two-sided report, produced in alphabetical sequence for students within the school. Two copies of this report are produced for every student tested, one for the student's permanent folder after the results are analyzed, and the other for the student's parent/guardian to be shared in a manner determined by the local district.

The scale scores in Language Arts Literacy and Mathematics are provided on the front of the ISR (Figure 10.1.1) of this report. There is also explanatory text here about scale scores and proficiency levels. Cluster data is provided on the back of the ISR (Figure 10.2.1) of this report. There is also explanatory text here about cluster scores.

The Just Proficient Mean is a statewide statistic comprised of the average or mean score attained on each cluster by all students (GE, SE, and 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.

The ISR for NJ ASK4 is shown in sample format as Figure 10.1.1 (front page) and Figure 10.1.2 (back page).

Figure 10.1.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Individual Student Report (ISR) – Front**

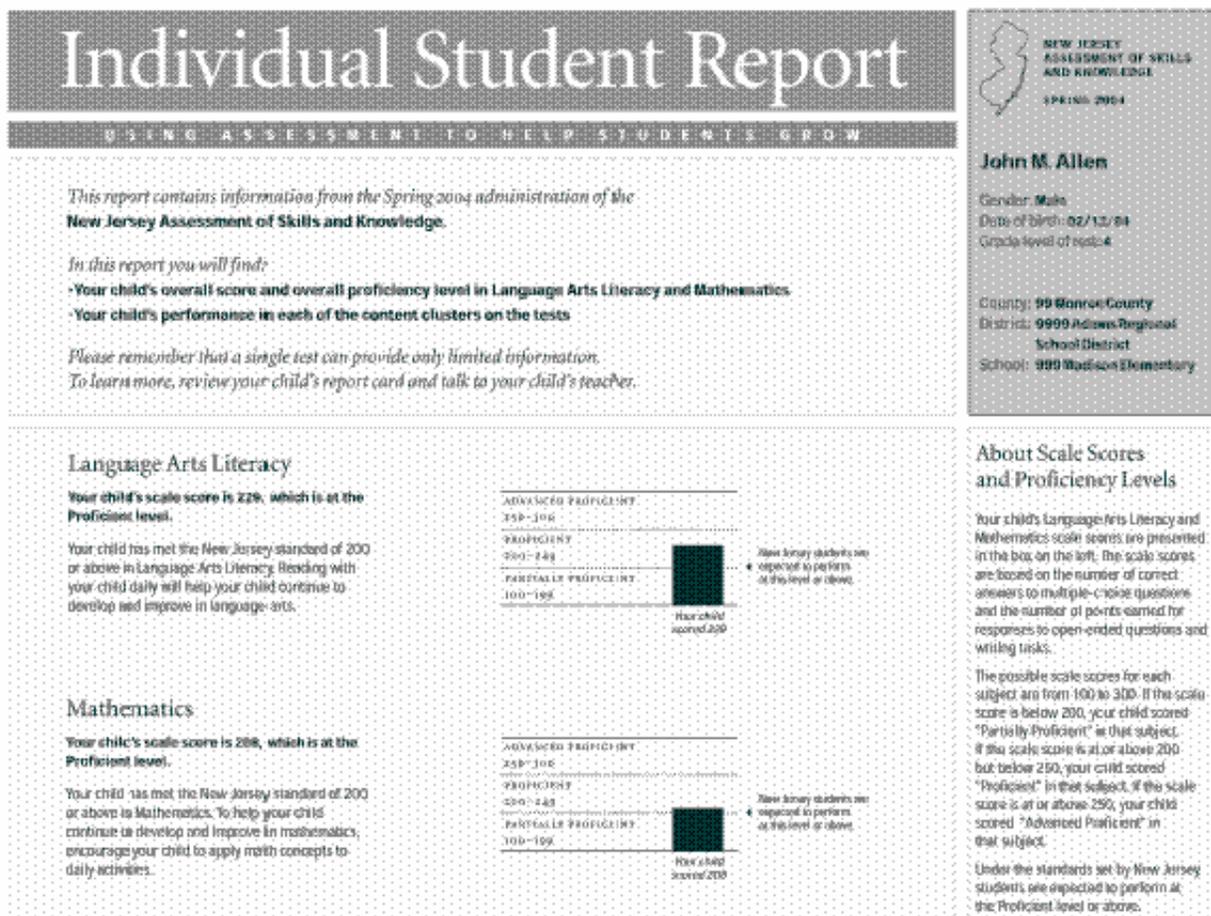
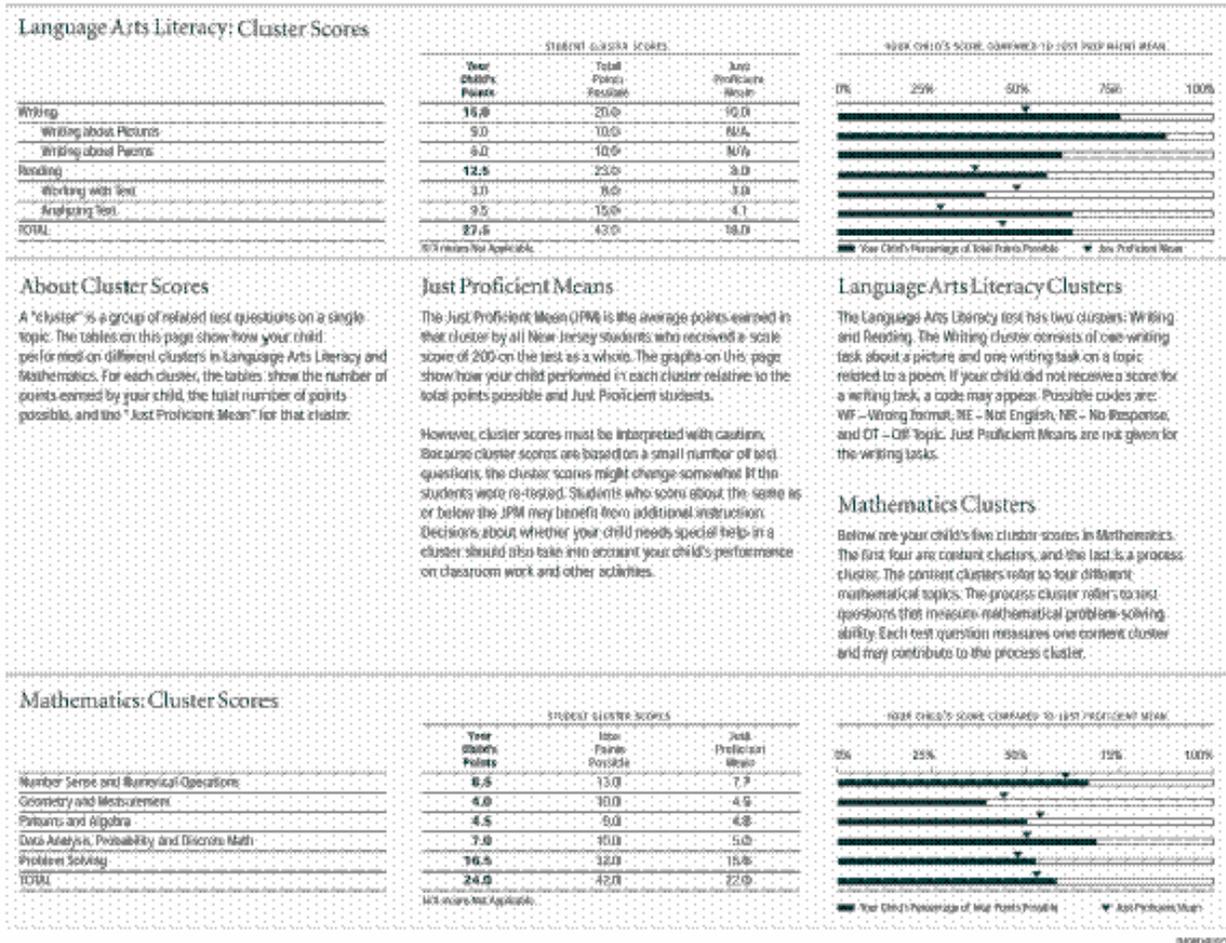


Figure 10.1.2

2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
Individual Student Report (ISR) – Back



All Sections Roster

The All Sections Roster 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.

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 will appear in this space.

Student Roster – Language Arts Literacy

The Student Roster – Language Arts Literacy lists the names of the students (last name first) in groups by proficiency level. Thus, the first students listed on the Language Arts Literacy roster are the students with the highest Language Arts Literacy scale scores. Students are listed alphabetically when more than one student has earned the same score. Students whose test booklets were voided and students coded APA, who are exempt from taking the test, are listed alphabetically at the end of the roster.

Following a student’s identification information, the student’s Language Arts Literacy scale score is given. This score is based on a combination of the number of correct answers to multiple-choice items and the number of points earned for open-ended items and writing tasks. Points earned are then reported for each cluster. Each item contributes only once to the NJ ASK total score.

Student Roster – Mathematics

The Student Roster – Mathematics lists the names of the students (last name first) in groups by proficiency level. Thus, the first students listed on the Mathematics roster are the students with the highest Mathematics scale scores. Students are listed alphabetically when more than one student has achieved the same score. Students whose test booklets were voided and students coded APA, who are exempt from taking the test, are listed alphabetically at the end of the roster.

Following a student’s identification information, the student’s total Mathematics score is given. This score is based on a combination of the number of correct answers to multiple-choice items and the number of points earned for open-ended items. Points earned are then reported for each cluster. Each item contributes only once to the NJ ASK total score.

Summary of School Performance

There are two Summary of School Performance reports, one for Language Arts Literacy and one for Mathematics. The reports are produced at the school level and provide preliminary aggregated data for a test section. Final aggregated data is sent in Cycle II. Data are provided for total students, general education students, special education students, and limited English proficient students. Data are also presented in the report by gender, ethnicity, economic status, and migrant status.

The report provides the percent of students in each proficiency level as well as the number of total students, general education students, special education students, and limited English proficient students tested for each content area.

Summary of District Performance

There are two Summary of District Performance reports, one for Language Arts Literacy and one for Mathematics which provide aggregated data for the district. In addition, this report includes data for total students, general education students, special education students, and limited English proficient students combined. The report format is the same as the summary of school performance. Any district that chooses to test a student classified Alternate Proficiency Assessment (APA), who is exempt from taking the NJ ASK, will receive score reports for that student, and the scores will be aggregated into the school and district reports.

Summary of School Cluster Performance

There are two Summary of School Cluster Performance reports, one for Language Arts Literacy and one for Mathematics. The reports are produced at the school level and provide aggregated data for each test section. Data are provided for general education students, special education students, limited English proficient students, and Title I students. Cluster level means for each of these populations are also presented on this report.

Summary of District Cluster Performance

There are two Summary of District Cluster Performance reports; one for Language Arts Literacy and one for Mathematics, which provide aggregated data for the district. In addition, this report includes data for total students, general education students, special education students, limited English proficient students, and Title I students combined. The report format is the same as the summary of school cluster performance. Any district that chooses to test a student classified Alternate Proficiency Assessment (APA), who is exempt from taking the NJ ASK, will receive score reports for that student, and the scores will be aggregated into the school and district reports.

10.2 Cycle II Reports

The Cycle II reports include the following: School and District Reports, Special School Reports, Statewide Disaggregated Student Population Report, DFG Report, Special Needs and Non-Special Needs Report, and Title I Report. Each Cycle II report is briefly described below.

School and District Reports

The school and district reports provide a complete analysis of student performance. Separate reports are produced for each subject tested. Each report covers two pages. The first page of each report provides information pertaining to total students, general education students, special education students, and limited English proficient students, as well as to groups classified by gender, ethnicity, economic status, and migrant status. The second page provides cluster raw score information for total students, general education students, special education students, limited English proficient students, and Title I students.

District/Schools identified as “Special Needs” have additional data. Special Needs District Mean, as calculated for total students, statewide, in a district identified as “Special Needs”. Non-Special Needs District Mean, as calculated for total students, statewide, in a district not identified as “Special Needs”.

The School Report for NJ ASK4 is shown in sample format as Figure 10.2.1 (front page – Performance by Demographic Groups) and Figure 10.2.2 (back page – Cluster Score Means).

Figure 10.2.1

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
School Report - Performance by Demographic Groups**

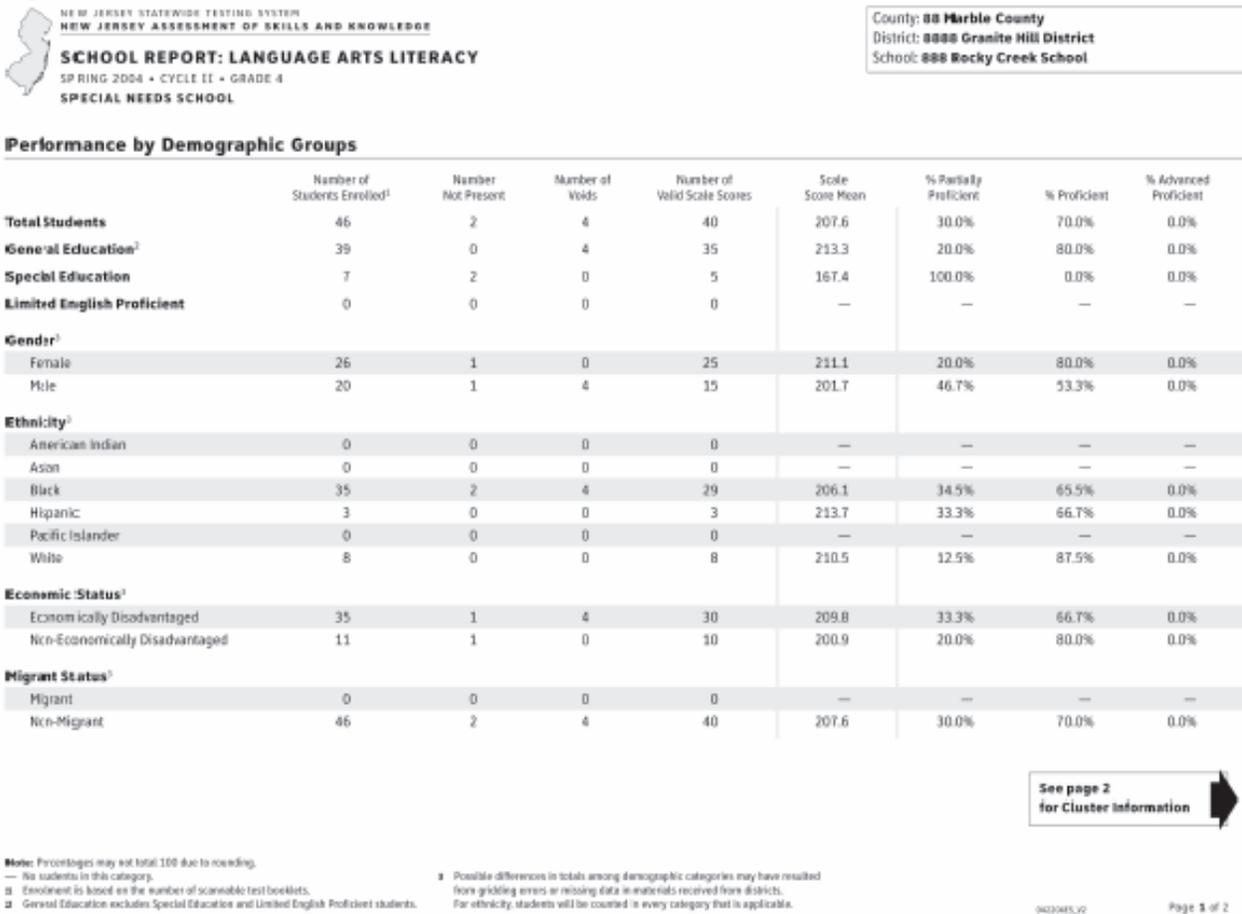


Figure 10.2.2

**2004 New Jersey Assessment of Skills and Knowledge (NJ ASK)
School Report - Cluster Score Means**

SCHOOL REPORT: LANGUAGE ARTS LITERACY
SPRING 2004 • CYCLE II • GRADE 4
SPECIAL NEEDS SCHOOL

County: **08 Marble County**
District: **0000 Granite Hill District**
School: **088 Rocky Creek School**

Cluster Score Means¹

	Writing	Writing About Pictures	Writing About Poems	Reading	Working with Text	Analyzing Text	TOTAL
Total Points Possible	20.0	10.0	10.0	20.0	10.0	15.0	43.0
Just Proficient Means ²	10.0	5.1	4.8	8.0	3.9	4.1	18.0
Total Students	10.2	5.0	5.2	10.6	5.0	5.6	20.8
General Education Students ³	10.7	5.3	5.4	11.6	5.5	6.1	22.3
Special Education Students	6.6	2.6	4.0	3.3	1.8	1.5	9.9
Limited English Proficient Students	—	—	—	—	—	—	—
Title I Students	11.5	5.3	6.3	11.0	5.3	5.8	22.5
District Mean	10.6	5.6	5.0	10.6	4.9	5.7	21.2
DFG B Mean	10.9	5.6	5.3	10.1	4.7	5.4	21.0
Special Needs District Mean	11.1	5.7	5.4	10.8	4.9	6.0	21.9
Non-Special Needs District Mean	11.9	6.1	5.8	11.6	5.3	6.3	23.5
State Mean	11.7	6.0	5.7	11.4	5.2	6.3	23.1

Students coded both SE and LEP 0
Students coded ANA who did not take the NJ ASK 2
Students coded void 4

- 1. Cluster means exclude students who took large-print, Braille, and alternate forms, as well as students with voided test booklets.
- 2. Just Proficient Means are the statewide raw score means for students whose scale score is 200.
- 3. General Education excludes Special Education and Limited English Proficient students.

04/20/05,NJ

Page 2 of 2

Special School Reports

Special reports are produced where a district requests information about the performance of special groups, as identified by the district at the time of testing. By using the “special” code category at the time of the test administration, districts have the opportunity to create such reports for specific student groups containing six or more students. Student test booklets may be coded in any of the four two-column “Special Codes” grids labeled A, B, C, and D. The special code, as coded on the students’ test booklet, is printed in the report title. Special reports are produced at the school level. One report for each content area per code is produced.

Statewide Disaggregated Student Population Report

Statewide Disaggregated Student Population Reports provide state-level data pertaining to the performance of the population groups identified below. Separate reports are provided for each subject tested.

DFG Report

The DFG Report summarizes the performance data for the DFG into which the district receiving the report has been categorized. Results are reported by total students, general education students, special education students and limited English proficient students as well as to groups classified by gender, ethnicity, economic status and migrant status.

Special Needs And Non-Special Needs Report

The Special Needs and Non-Special Needs Reports are distributed to New Jersey special needs districts. These reports summarize the statewide performance of special needs districts and the statewide performance of non-special needs districts. Results are reported by total students, general education students, special education students and limited English proficient students as well as to groups classified by gender, ethnicity, economic status and migrant status.

Title I Report

Title I Reports summarize the performance of Title I students statewide. Results are reported by total students, general education students, special education students and limited English proficient students as well as to groups classified by gender, ethnicity, economic status and migrant status.

10.3 State Summary Reporting

The State Summary consists of a group of files presented to the State on a CD. These files include an executive summary, report PDFs, and test result tables and graphs.

The executive summary contains a brief history of each test and the highlights of 2004 results based on the state Cycle II demographic report. The executive summaries for Grades 3 and 4 can be found in Appendix A. Additional statewide Cycle II results can be found in Appendix B. Two files of test results are provided based on the Cycle II demographic reports; one file includes all data with no suppression rules applied, and the other file applies the suppression rules for small cell numbers. The suppression rules are included in the executive summaries in Appendix A.

The report PDFs included on the State Summary CD are the Cycle II Performance Reports by Demographic Groups, the DFG Reports (except DFGs O and S), and the Special Needs and Non-Special Needs Reports.

Longitudinal data graphs of percentages proficient and above for demographic groups from the first time each test was administered to 2004 are also provided in the State Summary.

10.4 Interpreting Reports

The 2004 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 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 subjects tested. For each subject, 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.

10.5 Quality Control in Reporting

Prior to reports being distributed, both the reports themselves and the steps leading up to the production of the reports are subjected to extensive quality control procedures. These procedures include tasks to ensure the raw scores are accurately recorded in the database, and to ensure the scale scores and proficiency levels have been converted accurately. The aggregated data file is extensively reviewed to ensure the data is aggregated according to the aggregation rules defined by the State. The paper reports are then reviewed to verify all of the data is accurately represented on each report.

APPENDIX A:
Statewide Cycle II Executive Summary Results

Grade 3 New Jersey Assessment of Knowledge and Skills
Spring 2004

Executive Summary

The spring 2004 grade 3 New Jersey Assessment of Knowledge and Skills (NJ ASK) consisted of two content areas: Language Arts Literacy and Mathematics. 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 Core Curriculum Content Standards. 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 Core Curriculum Content Standards. 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 scores are reported as scale scores in each of the content areas. 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 NJ ASK was administered between March 16 and March 19, 2004. From a total third grade student population of 104,962, valid scores were obtained in language arts literacy from 103,414 students, with 968 students not present and 580 voids (unscorable due to illness, other difficulties during testing, or an insufficient number of items answered in a given content area). Valid scores were obtained in mathematics from 103,559 students, with 1150 not present and 253 voids. Performance levels for the grade 3 NJ ASK tests were established by panels of educators during standard-setting sessions held between June 28 and July 6, 2004. The grade 3 standards were approved by the New Jersey State Board of Education on July 7, 2004.

This executive summary includes two tables summarizing statewide test results for the 2004 grade 3 administration of the NJ ASK. Table A.3.1 presents results for Language Arts Literacy and Table A.3.2 presents results for Mathematics. Results are presented for the following student groups: total, general education, special education, and limited English proficient students. Data are also summarized for several demographic variables including: gender, ethnicity, and economic status. The tables include the number of students enrolled, not present, voided, and with valid scale scores. In addition, the tables present mean scale scores and the percent of students in each performance category (i.e., Partially Proficient, Proficient, and Advanced Proficient).

The tables that follow are derived from the statewide performance data of the Cycle II report. Note that the enrollment is based on the number of scannable test booklets, and students coded multiple ethnicity are counted in every ethnic category that is applicable. The percentage of students in Proficient or Advanced Proficient is calculated by subtracting the percentage of students in Partially Proficient from one hundred. The percentages may not total to one hundred due to rounding.

Following are highlights of the 2004 third grade assessment results.

Grade 3 Language Arts Literacy Results:

- Of the 103,414 grade 3 students with valid scale scores in Language Arts Literacy in Spring 2004, 20.7% scored in Partially Proficient; 75.6% scored in Proficient and 3.8% scored in Advanced Proficient (Table A.3.1).
- **General Education** 86.4% of general education grade 3 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.3.1).
- **Special Education** 50.1% of special education grade 3 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.3.1).
- **Limited English Proficient** 48.7% of limited English proficient grade 3 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.3.1).
- **Gender** 83.5% of female compared to 75.4% of male grade 3 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.3.1).
- **Ethnicity** For performance by ethnicity groups, students scoring in Proficient or Advanced Proficient in Language Arts Literacy ranged from 92.6% of Pacific Islanders to 61.8% of African American students. The percentage of Proficient and Advanced Proficient for all other race/ethnic groups fell between Pacific Islanders and African Americans (Table A.3.1).
- **Economic Status** 61.3% of economically disadvantaged grade 3 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.3.1).
- The mean scale score for all grade 3 students on the Language Arts Literacy test in Spring 2004 was 215.5 (Table A.3.1).

Grade 3 Mathematics Results:

- Of the 103,559 grade 3 students with valid scale scores in Mathematics in Spring 2004, 23.4% scored in Partially Proficient; 53.8% scored in Proficient and 22.8% scored in Advanced Proficient (Table A.3.2).

- **General Education** 81.6% of general education grade 3 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.3.2).
- **Special Education** 56.2% of special education grade 3 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.3.2).
- **Limited English Proficient** 54.8% of limited English proficient grade 3 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.3.2).
- **Gender** 77.3% of female compared to 75.9% of male grade 3 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.3.2).
- **Ethnicity** For performance by ethnicity groups, students scoring in Proficient or Advanced Proficient in Mathematics ranged from 90% of Asian American students to 54% of African American students. The percentage of Proficient and Advanced Proficient for all other race/ethnic groups fell between Pacific Islanders and African Americans (Table A.3.2).
- **Economic Status** 58.1% of economically disadvantaged grade 3 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.3.2).
- The mean scale score for all grade 3 students on the Mathematics test in Spring 2004 was 222.2 (Table A.3.2).

Reporting Rules for Data File:

The accompanying state summary data file contains the same type of information shown in the statewide summary tables included with this executive summary. Please note that there may be small differences between the state summary data file and the Cycle II reports issued to districts. In order to safeguard student confidentiality, certain information is suppressed in the state summary file according to the following reporting rules:

- Data are not reported where the number of students with valid scale scores for a particular group is less than 11.
- Data are not reported where demographic groups are mutually exclusive (e.g., gender) and there are one or two students with a valid scale score in one of the groups (e.g., male).
- Data are not reported when it is otherwise possible to identify individual student performance.

TABLE A.3.1

STATEWIDE PERFORMANCE BY DEMOGRAPHIC GROUPS SPRING 2004 -- GRADE 3

Language Arts Literacy

	Number of Students Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	Scale Score Mean	% Partially Proficient	% Proficient	% Advanced Proficient
Total Students	104962	968	580	103414	215.5	20.7%	75.6%	3.8%
General Education	83737	272	303	83162	220.1	13.6%	81.9%	4.5%
Special Education	15411	634	213	14564	196.7	49.9%	49.3%	0.8%
Limited English Proficient	6334	71	72	6191	194.8	51.3%	48.0%	0.8%
Gender								
Female	50753	339	189	50225	219.3	16.5%	77.9%	5.6%
Male	54117	582	385	53150	212.0	24.6%	73.4%	2.1%
Ethnicity								
American Indian	203	1	2	200	212.2	26.5%	72.0%	1.5%
Asian	7420	47	28	7345	225.0	9.5%	82.3%	8.2%
Black	18779	236	203	18340	204.0	38.2%	60.8%	1.1%
Hispanic	18898	237	145	18516	204.6	35.2%	63.9%	0.9%
Pacific Islander	699	6	4	689	225.2	7.4%	83.2%	9.4%
White	58749	346	189	58214	221.4	12.0%	83.0%	5.0%
Economic Status								
Economically Disadvantaged	31428	368	299	30761	203.1	38.7%	60.4%	0.9%
Non-Economically Disadvantaged	73534	600	281	72653	220.8	13.1%	82.0%	5.0%

TABLE A.3.2

STATEWIDE PERFORMANCE BY DEMOGRAPHIC GROUPS SPRING 2004 -- GRADE 3

Mathematics

	Number of Students Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	Scale Score Mean	% Partially Proficient	% Proficient	% Advanced Proficient
All Students	104962	1150	253	103559	222.2	23.4%	53.8%	22.8
Education Students	83737	411	133	83193	226.5	18.4%	55.7%	25.9
Education Students	15411	691	95	14625	204.7	43.8%	45.7%	10.5
High Proficient Students	6334	51	28	6255	203.4	45.2%	45.8%	9.0%
Gender								
Female	50753	424	83	50246	222.9	22.7%	53.9%	23.4
Male	54117	711	161	53245	221.6	24.1%	53.7%	22.3
Ethnicity								
American Indian	203	0	1	202	219.5	25.7%	53.5%	20.8
Asian	7420	44	12	7364	237.8	10.0%	48.2%	41.8
Hispanic	18779	296	74	18409	202.7	46.0%	46.1%	7.9%
Latino	18898	248	65	18585	209.8	36.5%	52.1%	11.4
Native Hawaiian or Other Pacific Islander	699	8	1	690	237.4	10.7%	47.7%	41.6
White	58749	496	83	58170	230.3	13.9%	57.4%	28.8
Economic Status								
Economically Disadvantaged	31428	428	109	30891	205.8	41.9%	48.7%	9.4%
Non-Economically Disadvantaged	73534	722	144	72668	229.2	15.6%	55.9%	28.5

Grade 4 New Jersey Assessment of Knowledge and Skills Spring, 2004

Executive Summary

The spring 2004 grade 4 New Jersey Assessment of Knowledge and Skills (NJ ASK) consisted of two content areas: Language Arts Literacy and Mathematics. Science was administered as a field test to grade four students in 2004. Thus, science results will not be presented here. 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 Core Curriculum Content Standards. 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 Core Curriculum Content Standards. 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 scores are reported as scale scores in each of the content areas. 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 NJ ASK was administered between March 16 and March 19, 2004. From a total fourth grade student population of 105,340, valid scores were obtained in language arts literacy from 103,818 students, with 869 students not present and 653 voids (unscorable due to illness, other difficulties during testing, or an insufficient number of items answered in a given content area.) Valid scores were obtained in mathematics from 103,770 students, with 899 not present and 671 voids.

This executive summary includes two tables summarizing statewide test results for the 2004 administration of the grade 4 NJ ASK. Table A.4.1 presents results for Language Arts Literacy and Table A.4.2 presents results for Mathematics. Results are presented for the following student groups: total, general education, special education, and limited English proficient students. Data are also summarized for several demographic variables including: gender, ethnicity, and economic status. The tables include the number of students enrolled, not present, voided, and with valid scale scores. In addition, the tables present mean scale scores and the percentage of students in each performance category (i.e., Partially Proficient, Proficient, and Advanced Proficient).

The tables that follow are derived from the statewide performance data of the Cycle II report. As a result of adjustments made to the data by districts after the reporting deadline, small differences may exist between the information in these tables and the reports issued to districts. Note that the enrollment is based on the number of scannable test booklets, and students coded multiple ethnicity are counted in every ethnic category that is applicable. The percentage of students in Proficient or Advanced Proficient is calculated by subtracting the percentage of

students in Partially Proficient from one hundred. The percentages may not total to one hundred due to rounding.

Following is a list of highlights from the two tables included in the Executive Summary.

Grade 4 Language Arts Literacy Results:

- Of the 103,818 grade 4 students with valid scale scores in Language Arts Literacy in spring, 2004, 17.9% scored in Partially Proficient; 77.5% scored in Proficient and 4.7% scored in Advanced Proficient (Table A.4.1).
- **General Education:** 90.3% of general education grade 4 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.4.1).
- **Special Education:** 49.0% of special education grade 4 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.4.1).
- **Limited English Proficient:** 48.6% of limited English proficient grade 4 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.4.1).
- **Gender:** 86.6% of female compared to 77.9% of male grade 4 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.4.1).
- **Ethnicity:** For performance by ethnic groups, students scoring in Proficient or Advanced Proficient in Language Arts Literacy ranged from 91.7% of Pacific Islanders to 66.8% African Americans. The percentage of Advanced Proficient and Proficient for all other race/ethnic groups fell between Pacific Islanders and African Americans (Table A.4.1).
- **Economic Status:** 66.2% of economically disadvantaged grade 4 students in 2004 scored in Proficient or Advanced Proficient in Language Arts Literacy (Table A.4.1).
- The mean scale score for all grade 4 students on the Language Arts Literacy test in spring, 2004, was 217.7 (Table A.4.1).

Grade 4 Mathematics Results:

- Of the 103,770 grade 4 students with valid scale scores in Mathematics in spring, 2004, 27.9% scored in Partially Proficient; 46.4% scored in Proficient and 25.7% scored in Advanced Proficient (Table A.4.2).
- **General Education:** 78.4% of general education grade 4 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.4.2).

- **Special Education:** 46.3% of special education grade 4 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.4.2).
- **Limited English Proficient:** 47.2% of limited English proficient grade 4 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.4.2).
- **Gender:** 71.7% of female compared to 72.5% of male grade 4 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.4.2).
- **Ethnicity:** For performance by ethnic groups, students scoring in Proficient or Advanced Proficient in Mathematics ranged from 87.8% of Asians to 50.2% African Americans. The percentage of Advanced Proficient and Proficient for all other race/ethnic groups fell between Asians and African Americans (Table A.4.2).
- **Economic Status:** 54.1% of economically disadvantaged grade 4 students in 2004 scored in Proficient or Advanced Proficient in Mathematics (Table A.4.2).
- The mean scale score for all grade 4 students on the Mathematics test in spring 2004 was 221.4 (Table A.4.2).

Reporting Rules for Data File

The accompanying state summary data file contains the same type of information shown in the statewide summary tables included with this executive summary. Please note that there may be small discrepancies differences between the data file and reports issued to districts due to adjustments made to the data by districts after the reporting deadline. In order to safeguard student confidentiality, certain information is suppressed in the state summary file according to the following reporting rules:

- Data are not reported where the number of students with valid scale scores for a particular group is less than 11.
- Data are not reported where demographic groups are mutually exclusive (e.g., gender) and there are one or two students with a valid scale score in one of the groups (e.g., male).
- Data are not reported when it is otherwise possible to identify individual student performance.

TABLE A.4.1

STATEWIDE PERFORMANCE BY DEMOGRAPHIC GROUPS SPRING 2004 -- GRADE 4

Language Arts Literacy

	Number of Students Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	Scale Score Mean	% Partially Proficient	% Proficient	% Advanced Proficient
Total Students	105340	869	653	103818	217.7	17.9%	77.5%	4.7%
General Education	83617	184	289	83144	223.2	9.7%	84.6%	5.7%
Special Education	16726	645	287	15794	195.7	51.0%	48.4%	0.7%
Limited English Proficient	5443	49	88	5306	192.7	51.4%	47.9%	0.8%
Gender								
Female	51181	273	239	50669	221.7	13.4%	79.8%	6.7%
Male	54093	563	409	53121	213.8	22.1%	75.2%	2.7%
Ethnicity								
American Indian	235	2	0	233	215.9	19.7%	76.0%	4.3%
Asian	7316	47	25	7244	227.3	8.4%	81.2%	10.4%
Black	19268	257	255	18756	205.9	33.2%	65.7%	1.1%
Hispanic	18293	169	189	17935	206.4	31.0%	67.8%	1.2%
Pacific Islander	708	2	4	702	227.3	8.3%	81.6%	10.1%
White	59318	312	173	58833	223.6	10.2%	83.7%	6.1%
Economic Status								
Economically Disadvantaged	31462	320	393	30749	204.8	33.8%	65.4%	0.8%
Non-Economically Disadvantaged	73878	549	260	73069	223.1	11.2%	82.5%	6.3%

TABLE A.4.2

STATEWIDE PERFORMANCE BY DEMOGRAPHIC GROUPS SPRING 2004 -- GRADE 4

Mathematics

	Number of Students Enrolled	Number Not Present	Number of Voids	Number of Valid Scale Scores	Scale Score Mean	% Partially Proficient	% Proficient	% Advanced Proficient
	105340	899	671	103770	221.4	27.9%	46.4%	25
Education	83617	195	390	83032	227.0	21.6%	48.9%	29
Proficient	16726	673	241	15812	198.2	53.7%	36.4%	10
Advanced Proficient	5443	40	42	5361	200.1	52.8%	36.0%	11
	105340	899	671	103770	221.4	27.9%	46.4%	25
Gender								
Male	51181	299	276	50606	220.6	28.3%	47.6%	24
Female	54093	582	387	53124	222.1	27.5%	45.3%	27
Ethnicity								
American Indian	235	2	0	233	222.1	27.5%	45.5%	27
Asian	7316	44	29	7243	239.9	12.2%	40.9%	46
Hispanic	19268	283	245	18740	202.0	49.8%	39.6%	10
Other	18293	181	136	17976	209.2	40.8%	44.4%	14
Native Hawaiian or Other Pacific Islander	708	2	3	703	238.7	14.1%	38.8%	47
White	59318	331	250	58737	228.9	19.0%	49.9%	31
Economic Status								
Economically Disadvantaged	31462	366	336	30760	205.2	45.9%	41.7%	12
Non-Economically Disadvantaged	73878	533	335	73010	228.1	20.3%	48.4%	31

**APPENDIX B:
Additional Statewide Cycle II Results**

TABLE B.3.1

**NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004
NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE BY DISTRICT
FACTOR GROUP**

LANGUAGE ARTS LITERACY SECTION – Grade 3

GENERAL EDUCATION STUDENTS^b

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	14,080	66.2	33.8	65.2	1.0	206.4
B	8,479	81.1	18.9	79.5	1.5	214.9
CD	7,077	83.9	16.1	81.4	2.5	216.7
DE	12,707	90.3	9.7	86.7	3.5	221.3
FG	10,313	91.8	8.2	87.6	4.3	223.1
GH	11,637	93.3	6.7	87.4	5.9	225.1
I	16,249	96.5	3.5	87.1	9.4	229.5
J	1,530	96.8	3.2	86.3	10.5	230.6

SPECIAL EDUCATION STUDENTS^c

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	2,293	24.9	75.1	24.9	0.1	180.6
B	1,705	37.5	62.5	37.3	0.2	190.3
CD	1,387	42.3	57.7	41.9	0.4	192.2
DE	2,143	49.4	50.6	48.8	0.6	196.6
FG	1,878	56.0	44.0	55.0	1.0	200.9
GH	2,026	60.1	39.9	58.9	1.2	202.8
I	2,751	72.6	27.4	70.8	1.8	209.4
J	221	66.5	33.5	65.6	0.9	206.0

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.
b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.
c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.
d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.
e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.3.1 (continued)

**NEW JERSEY STATEWIDE TESTING SYSTEM
SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE
BY DISTRICT FACTOR GROUP**

LANGUAGE ARTS LITERACY SECTION – Grade 3

LIMITED ENGLISH PROFICIENT STUDENTS ^d

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	3,561	43.4	56.6	43.0	0.4	191.6
B	630	39.5	60.5	39.0	0.5	188.2
CD	335	49.0	51.0	48.7	0.3	195.4
DE	464	61.0	39.0	60.1	0.9	201.4
FG	381	62.5	37.5	60.6	1.8	202.2
GH	405	63.2	36.8	62.0	1.2	204.4
I	392	69.6	30.4	66.6	3.1	209.7
J	14	57.1	42.9	57.1	0.0	201.6

TOTAL STUDENTS ^e

DFG	NUMBER TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	19,666	57.8	42.2	57.0	0.8	201.1
B	10,766	72.0	28.0	70.8	1.3	209.6
CD	8,774	76.2	23.8	74.1	2.1	212.2
DE	15,281	83.8	16.2	80.8	3.0	217.3
FG	12,542	85.7	14.3	82.0	3.7	219.2
GH	14,018	87.8	12.2	82.7	5.1	221.4
I	19,352	92.6	7.4	84.4	8.2	226.4
J	1,765	92.7	7.3	83.5	9.2	227.3

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.
- b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.
- c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.
- d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.
- e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.3.1 (continued)

**NEW JERSEY STATEWIDE TESTING SYSTEM
SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE**

LANGUAGE ARTS LITERACY SECTION – Grade 3

CHARTER SCHOOLS^f

	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	1,083	70.7	29.3	68.9	1.8	210.0
SPECIAL ^c EDUCATION STUDENTS	99	26.3	73.7	26.3	0.0	187.0
LIMITED ENGLISH ^d PROFICIENT STUDENTS	1	100	0.0	100	0.0	204.0
TOTAL ^e STUDENTS	1,182	67.0	33.0	65.3	1.7	208.1

STATEWIDE RESULTS

	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	83,162	86.4	13.6	81.9	4.5	220.1
SPECIAL ^c EDUCATION STUDENTS	14,564	50.1	49.9	49.3	0.8	196.7
LIMITED ENGLISH ^d PROFICIENT STUDENTS	6,191	48.7	51.3	48.0	0.8	194.8
TOTAL ^e STUDENTS	103,414	79.3	20.7	75.6	3.8	215.5

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.

b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.

c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.

d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.

e. INCLUDES ALL STUDENTS TESTED.

f. CHARTER SCHOOLS ARE NOT INCLUDED IN A DFG.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.3.2

**NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004
NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE BY DISTRICT
FACTOR GROUP**

MATHEMATICS SECTION – Grade 3

GENERAL EDUCATION STUDENTS ^b

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	14,132	57.8	42.2	47.7	10.1	206.1
B	8,491	75.2	24.8	57.8	17.4	219.3
CD	7,061	80.4	19.6	60.7	19.7	223.5
DE	12,718	85.9	14.1	60.3	25.7	229.0
FG	10,327	86.9	13.1	59.6	27.3	230.3
GH	11,661	90.1	9.9	56.9	33.2	234.6
I	16,178	93.6	6.4	53.7	39.9	239.1
J	1,531	93.8	6.2	49.3	44.5	240.6

SPECIAL EDUCATION STUDENTS ^c

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	2,335	33.2	66.8	29.8	3.4	186.6
B	1,714	47.2	52.8	41.0	6.2	197.6
CD	1,389	51.6	48.4	43.8	7.8	200.9
DE	2,145	56.4	43.6	46.7	9.7	204.8
FG	1,888	60.4	39.6	49.2	11.2	208.1
GH	2,033	67.4	32.6	53.5	14.0	213.7
I	2,739	72.6	27.4	54.7	17.9	217.5
J	221	69.2	30.8	53.8	15.4	214.4

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.

b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.

c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.

d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.

e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.3.2 (continued)

**NEW JERSEY STATEWIDE TESTING SYSTEM
 SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE
 BY DISTRICT FACTOR GROUP**

MATHEMATICS SECTION – Grade 3

LIMITED ENGLISH PROFICIENT STUDENTS ^d

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	3,585	51.4	48.6	43.9	7.4	201.1
B	635	44.4	55.6	39.4	5.0	193.9
CD	340	58.5	41.5	53.8	4.7	203.3
DE	467	64.7	35.3	53.3	11.3	209.1
FG	391	58.8	41.2	45.5	13.3	208.0
GH	414	65.2	34.8	51.2	14.0	211.6
I	399	71.9	28.1	52.1	19.8	219.7
J	15	73.3	26.7	33.3	40.0	223.0

TOTAL STUDENTS ^e

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	19,780	54.0	46.0	45.1	8.9	203.1
B	10,792	69.1	30.9	54.1	15.0	214.5
CD	8,765	75.1	24.9	57.8	17.3	219.2
DE	15,297	81.3	18.7	58.2	23.1	225.1
FG	12,573	82.2	17.8	57.6	24.5	226.4
GH	14,056	86.2	13.8	56.3	30.0	231.0
I	19,274	90.3	9.7	53.8	36.5	235.7
J	1,767	90.5	9.5	49.7	40.8	237.2

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.
- b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.
- c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.
- d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.
- e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.3.2 (continued)

**NEW JERSEY STATEWIDE TESTING SYSTEM
SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE**

MATHEMATICS SECTION – Grade 3

CHARTER SCHOOLS^f

	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	1,087	57.8	42.2	43.2	14.5	207.9
SPECIAL ^c EDUCATION STUDENTS	100	37.0	63.0	32.0	5.0	194.6
LIMITED ENGLISH ^d PROFICIENT STUDENTS	1	0.0	100	0.0	0.0	181.0
TOTAL ^e STUDENTS	1,187	56.0	44.0	42.3	13.7	206.8

STATEWIDE RESULTS

	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	83,193	81.6	18.4	55.7	25.9	226.5
SPECIAL ^c EDUCATION STUDENTS	14,625	56.2	43.8	45.7	10.5	204.7
LIMITED ENGLISH ^d PROFICIENT STUDENTS	6,255	54.8	45.2	45.8	9.0	203.4
TOTAL ^e STUDENTS	103,559	76.6	23.4	53.8	22.8	222.2

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.
- b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.
- c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.
- d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.
- e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.3.3

**NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004
NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE
LANGUAGE ARTS LITERACY AND MATHEMATICS
FOR THE SPECIAL NEEDS DISTRICTS AS COMPARED TO ALL OTHER DISTRICTS – Grade 3**

LANGUAGE ARTS LITERACY SECTION		NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
				PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	SPECIAL NEEDS	15,823	67.8	32.2	66.7	1.0	207.3
	ALL OTHERS	67,339	90.8	9.2	85.5	5.3	223.1
SPECIAL ^c EDUCATION STUDENTS	SPECIAL NEEDS	2,702	25.9	74.1	25.8	0.1	181.4
	ALL OTHERS	11,862	55.7	44.3	54.7	1.0	200.2
LIMITED ENGLISH ^d PROFICIENT STUDENTS	SPECIAL NEEDS	3,700	43.3	56.7	42.9	0.4	191.5
	ALL OTHERS	2,491	56.8	43.2	55.5	1.2	199.8
TOTAL ^e STUDENTS	SPECIAL NEEDS	21,942	59.1	40.9	58.2	0.8	201.8
	ALL OTHERS	81,472	84.8	15.2	80.2	4.6	219.2

MATHEMATICS SECTION		NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
				PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	SPECIAL NEEDS	15,874	59.7	40.3	49.2	10.5	207.4
	ALL OTHERS	67,319	86.7	13.3	57.2	29.5	231.0
SPECIAL ^c EDUCATION STUDENTS	SPECIAL NEEDS	2,746	33.9	66.1	30.4	3.5	187.1
	ALL OTHERS	11,879	61.3	38.7	49.2	12.1	208.7
LIMITED ENGLISH ^d PROFICIENT STUDENTS	SPECIAL NEEDS	3,722	51.0	49.0	43.7	7.2	200.8
	ALL OTHERS	2,533	60.4	39.6	48.8	11.6	207.3
TOTAL ^e STUDENTS	SPECIAL NEEDS	22,055	55.4	44.6	46.2	9.2	204.0
	ALL OTHERS	81,504	82.3	17.7	55.8	26.5	227.1

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.
b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.
c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.
d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.
e. INCLUDES ALL STUDENTS TESTED.
NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.4.1

**NEW JERSEY STATEWIDE TESTING SYSTEMS SPRING 2004
NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE BY DISTRICT
FACTOR GROUP**

LANGUAGE ARTS LITERACY SECTION – Grade 4

GENERAL EDUCATION STUDENTS^b

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	13,620	74.2	25.8	73.2	1.1	210.1
B	8,533	85.6	14.4	83.4	2.1	217.4
CD	7,123	89.4	10.6	86.4	3.0	220.6
DE	12,610	93.4	6.6	89.4	4.0	223.8
FG	10,282	94.6	5.4	88.7	5.9	226.2
GH	11,942	95.5	4.5	87.9	7.6	228.1
I	16,463	97.9	2.1	86.1	11.8	232.4
J	1,482	98.6	1.4	85.6	13.0	233.2

SPECIAL EDUCATION STUDENTS^c

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	2,780	23.0	77.0	22.9	0.1	178.8
B	1,879	38.4	61.6	38.2	0.2	189.9
CD	1,500	40.5	59.5	40.1	0.4	191.6
DE	2,325	51.1	48.9	50.7	0.4	197.2
FG	1,932	55.1	44.9	54.5	0.7	200.0
GH	2,154	59.5	40.5	58.5	1.0	202.2
I	2,820	72.2	27.8	70.8	1.4	209.2
J	229	73.4	26.6	70.7	2.6	209.8

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.

b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.

c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.

d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.

e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.4.1 (continued)

**NEW JERSEY STATEWIDE TESTING SYSTEM
SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE
BY DISTRICT FACTOR GROUP**

LANGUAGE ARTS LITERACY SECTION – Grade 4

LIMITED ENGLISH PROFICIENT STUDENTS ^d

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	3,110	44.6	55.4	44.1	0.5	189.6
B	531	44.6	55.4	44.3	0.4	189.6
CD	329	49.5	50.5	49.2	0.3	192.8
DE	348	52.6	47.4	51.1	1.4	197.9
FG	291	56.4	43.6	55.0	1.4	199.0
GH	340	60.9	39.1	60.3	0.6	201.9
I	334	66.5	33.5	63.8	2.7	205.4
J	20	80.0	20.0	75.0	5.0	218.1

TOTAL STUDENTS ^e

DFG	NUMBER TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	19,277	62.8	37.2	62.0	0.8	202.7
B	10,894	75.7	24.3	73.9	1.7	211.5
CD	8,931	79.9	20.1	77.5	2.5	214.8
DE	15,260	86.1	13.9	82.7	3.4	219.2
FG	12,484	87.6	12.4	82.7	5.0	221.5
GH	14,399	89.5	10.5	83.0	6.5	223.7
I	19,576	93.7	6.3	83.5	10.2	228.7
J	1,731	95.0	5.0	83.5	11.5	229.9

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.
- b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.
- c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.
- d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.
- e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.4.1 (continued)

**NEW JERSEY STATEWIDE TESTING SYSTEM
SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE**

LANGUAGE ARTS LITERACY SECTION – Grade 4

CHARTER SCHOOLS ^f

	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	1,086	72.4	27.6	70.8	1.6	209.7
SPECIAL ^c EDUCATION STUDENTS	122	23.0	77.0	23.0	0.0	184.5
LIMITED ENGLISH ^d PROFICIENT STUDENTS	2	50.0	50.0	50.0	0.0	179.0
TOTAL ^e STUDENTS	1,210	67.4	32.6	66.0	1.4	207.1

STATEWIDE RESULTS

	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	83,144	90.3	9.7	84.6	5.7	223.2
SPECIAL ^c EDUCATION STUDENTS	15,794	49.0	51.0	48.4	0.7	195.7
LIMITED ENGLISH ^d PROFICIENT STUDENTS	5,306	48.6	51.4	47.9	0.8	192.7
TOTAL ^e STUDENTS	103,818	82.1	17.9	77.5	4.7	217.7

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.

b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.

c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.

d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.

e. INCLUDES ALL STUDENTS TESTED.

f. CHARTER SCHOOLS ARE NOT INCLUDED IN A DFG.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.4.2

**NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004
NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE BY DISTRICT FACTOR GROUP**

MATHEMATICS SECTION – Grade 4

GENERAL EDUCATION STUDENTS ^b

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	13,625	59.6	40.4	43.1	16.4	210.6
B	8,529	70.6	29.4	49.7	20.9	219.0
CD	7,103	75.3	24.7	52.1	23.2	222.7
DE	12,578	80.7	19.3	52.7	28.0	227.9
FG	10,260	82.6	17.4	50.7	31.9	230.3
GH	11,916	85.4	14.6	50.5	35.0	233.4
I	16,451	90.5	9.5	47.3	43.2	239.3
J	1,481	91.7	8.3	49.4	42.3	240.1

SPECIAL EDUCATION STUDENTS ^c

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	2,799	27.2	72.8	23.4	3.8	181.7
B	1,886	39.0	61.0	32.4	6.6	191.8
CD	1,501	41.6	58.4	33.8	7.9	194.1
DE	2,329	48.4	51.6	38.9	9.5	199.9
FG	1,927	51.3	48.7	39.2	12.1	202.5
GH	2,155	52.5	47.5	40.4	12.1	203.9
I	2,815	63.0	37.0	46.4	16.6	212.1
J	229	63.8	36.2	46.7	17.0	213.3

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.

b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.

c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.

d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.

e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.4.2 (continued)

**NEW JERSEY STATEWIDE TESTING SYSTEM
 SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE
 BY DISTRICT FACTOR GROUP**

MATHEMATICS SECTION – Grade 4

LIMITED ENGLISH PROFICIENT STUDENTS ^d

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	3,146	44.7	55.3	34.5	10.2	198.0
B	533	39.2	60.8	31.9	7.3	194.8
CD	327	45.0	55.0	37.6	7.3	197.8
DE	352	49.4	50.6	39.5	9.9	202.2
FG	298	54.4	45.6	40.6	13.8	204.8
GH	347	55.9	44.1	41.2	14.7	207.1
I	334	65.3	34.7	40.7	24.6	216.4
J	21	81.0	19.0	57.1	23.8	223.0

TOTAL STUDENTS ^e

DFG	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	19,331	52.9	47.1	39.1	13.8	204.8
B	10,898	63.8	36.2	45.9	17.9	213.2
CD	8,909	68.6	31.4	48.6	20.1	217.1
DE	15,236	75.1	24.9	50.3	24.8	223.1
FG	12,464	77.2	22.8	48.7	28.4	225.5
GH	14,381	79.9	20.1	48.8	31.1	228.4
I	19,558	86.2	13.8	47.1	39.1	235.1
J	1,731	87.9	12.1	49.2	38.7	236.3

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.
- b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.
- c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.
- d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.
- e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.4.2 (continued)

**NEW JERSEY STATEWIDE TESTING SYSTEM
SPRING 2004 NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE**

MATHEMATICS SECTION – Grade 4

CHARTER SCHOOLS^f

	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	1,086	51.1	48.9	35.6	15.5	204.8
SPECIAL ^c EDUCATION STUDENTS	121	25.6	74.4	24.0	1.7	181.7
LIMITED ENGLISH ^d PROFICIENT STUDENTS	2	50.0	50.0	50.0	0.0	214.5
TOTAL ^e STUDENTS	1,209	48.6	51.4	34.5	14.1	202.5

STATEWIDE RESULTS

	NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	83,032	78.4	21.6	48.9	29.5	227.0
SPECIAL ^c EDUCATION STUDENTS	15,812	46.3	53.7	36.4	10.0	198.2
LIMITED ENGLISH ^d PROFICIENT STUDENTS	5,361	47.2	52.8	36.0	11.2	200.1
TOTAL ^e STUDENTS	103,770	72.1	27.9	46.4	25.7	221.4

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.
- b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.
- c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.
- d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.
- e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

TABLE B.4.3

**NEW JERSEY STATEWIDE TESTING SYSTEM SPRING 2004
NEW JERSEY ASSESSMENT OF SKILLS AND KNOWLEDGE
LANGUAGE ARTS LITERACY AND MATHEMATICS**

FOR THE SPECIAL NEEDS DISTRICTS AS COMPARED TO ALL OTHER DISTRICTS – Grade 4

LANGUAGE ARTS LITERACY SECTION		NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
				PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	SPECIAL NEEDS	15,389	75.4	24.6	74.3	1.1	210.8
	ALL OTHERS	67,755	93.6	6.4	87.0	6.7	226.0
SPECIAL ^c EDUCATION STUDENTS	SPECIAL NEEDS	3,225	25.3	74.7	25.2	0.1	180.2
	ALL OTHERS	12,569	55.1	44.9	54.3	0.8	199.7
LIMITED ENGLISH ^d PROFICIENT STUDENTS	SPECIAL NEEDS	3,226	45.2	54.8	44.6	0.5	190.2
	ALL OTHERS	2,080	53.9	46.1	52.8	1.1	196.6
TOTAL ^e STUDENTS	SPECIAL NEEDS	21,601	64.1	35.9	63.2	0.9	203.6
	ALL OTHERS	82,217	86.9	13.1	81.2	5.7	221.4

MATHEMATICS SECTION		NUMBER ^a TESTED 2004	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2004	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2004
				PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL ^b EDUCATION STUDENTS	SPECIAL NEEDS	15,399	60.6	39.4	44.0	16.7	211.4
	ALL OTHERS	67,633	82.5	17.5	50.0	32.5	230.5
SPECIAL ^c EDUCATION STUDENTS	SPECIAL NEEDS	3,250	28.3	71.7	24.6	3.7	182.4
	ALL OTHERS	12,562	51.0	49.0	39.4	11.6	202.2
LIMITED ENGLISH ^d PROFICIENT STUDENTS	SPECIAL NEEDS	3,265	45.1	54.9	34.8	10.4	198.6
	ALL OTHERS	2,096	50.3	49.7	38.0	12.4	202.5
TOTAL ^e STUDENTS	SPECIAL NEEDS	21,668	53.8	46.2	39.9	13.9	205.4
	ALL OTHERS	82,102	76.9	23.1	48.1	28.8	225.6

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND APA EXEMPT WITH NO SCALED SCORES.

b. EXCLUDES SPECIAL EDUCATION AND LIMITED ENGLISH PROFICIENT STUDENTS.

c. INCLUDES SPECIAL EDUCATION STUDENTS ONLY.

d. INCLUDES LIMITED ENGLISH PROFICIENT STUDENTS ONLY.

e. INCLUDES ALL STUDENTS TESTED.

NOTE: PERCENTAGES MAY NOT TOTAL 100 DUE TO ROUNDING

How to Interpret The Categories

The following is an explanation of how to interpret the categories of students presented in the following report. Please apply these rules as you read and interpret the report.

For each content area:

“General Education” excludes students coded as special education OR limited English proficient on their test booklets.

“Special Education” includes students coded as SE on their test booklet

“Limited English Proficient” includes students coded as LEP on their test booklet.

“Total” includes all students tested who were not Void.

DISTRICT FACTOR GROUPS

The District Factor Group (DFG) is an indicator of the socioeconomic status of citizens in each district and has been useful for the comparative reporting of test results from New Jersey's statewide testing programs. The measure was first developed in 1974 using demographic variables from the 1970 United States Census. A revision was made in 1984 to take into account new data from the 1980 United States Census. The DFG designations were updated again in 1992 using the following demographic variables from the 1990 United States Census.

- A. Percentage of adult residents who failed to complete high school
- B. Percentage of adult residents who attended college
- C. Occupational status of adult household members:
 - 1 = laborers
 - 2 = service workers (except private and protective)
 - 3 = farm workers
 - 4 = operatives and kindred workers
 - 5 = protective service workers
 - 6 = sales workers
 - 7 = clerical and kindred workers
 - 8 = craftsmen, foremen, and kindred workers
 - 9 = quasi-professionals
 - 10 = managers, officials, and proprietors
 - 11 = old and new professionals
- D. Population Density: persons per square mile
- E. Income: median family income
- F. Unemployment: percentage of those in the work force who received some unemployment compensation
- G. Poverty: percentage of residents below the poverty level

The variables described above were combined using a statistical technique called principal components analysis, which resulted in a single measure of socioeconomic status for each district. Districts were then ranked according to their score on this measure and divided into eight groups based on the score interval in which their scores were located. Eight DFGs have been created based on the 1990 United States Census data. They range from A (lowest socioeconomic districts) to J (highest socioeconomic districts) and are labeled as follows: A, B, CD, DE, FG, GH, I, J. Updating the DFGs has not changed any district's designation as Special Needs or not Special Needs.

Whereas the DFGs based on the 1980 United States Census resulted in 10 groups containing approximately equal numbers of districts, the DFGs based on the 1990 United States Census resulted in eight groups of different sizes depending on their score. The number of districts* in each DFG is now as follows:

DFG	Number of Districts
A	35
B	78
CD	75
DE	100
FG	87
GH	78
I	105
J	15

* Includes all New Jersey's public school districts (regardless of school configuration or grade levels served).

**APPENDIX C:
Raw to Scale Score Conversions**

Raw Score – Scale Score Conversions with Theta, S.E. and Cumulative Frequencies

TABLE C.3.1 Conversion 2004 NJ ASK Language Arts Literacy – Grade 3

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
0	128	-2.8523	1.392	13	0.0
0.5	130	-2.1844	0.726	15	0.0
1	132	-1.8312	0.494	79	0.1
1.5	134	-1.6380	0.392	82	0.1
2	136	-1.5074	0.333	211	0.2
2.5	138	-1.4095	0.294	221	0.2
3	140	-1.3309	0.267	397	0.4
3.5	142	-1.2648	0.247	413	0.4
4	144	-1.2074	0.232	647	0.6
4.5	146	-1.1560	0.221	680	0.7
5	148	-1.1091	0.212	952	0.9
5.5	150	-1.0655	0.205	1021	1.0
6	152	-1.0245	0.200	1367	1.3
6.5	154	-0.9852	0.196	1472	1.4
7	156	-0.9473	0.193	1909	1.8
7.5	158	-0.9105	0.191	2065	2.0
8	160	-0.8741	0.189	2556	2.5
8.5	162	-0.8382	0.189	2804	2.7
9	164	-0.8023	0.189	3371	3.3
9.5	166	-0.7664	0.189	3719	3.6
10	168	-0.7302	0.190	4345	4.2
10.5	170	-0.6935	0.192	4802	4.6
11	172	-0.6562	0.193	5597	5.4
11.5	174	-0.6182	0.195	6205	6.0
12	176	-0.5795	0.197	7140	6.9
12.5	178	-0.5399	0.200	7914	7.7
13	180	-0.4994	0.202	8976	8.7
13.5	182	-0.4579	0.204	9862	9.5
14	184	-0.4156	0.206	11163	10.8
14.5	186	-0.3724	0.209	12246	11.8
15	188	-0.3283	0.211	13676	13.2
15.5	190	-0.2833	0.212	15103	14.6
16	192	-0.2376	0.214	16872	16.3
16.5	194	-0.1911	0.216	18414	17.8
17	196	-0.1439	0.218	20455	19.8
17.5	198	-0.0961	0.219	21384	20.7
18	200	-0.0476	0.220	24085	23.3
18.5	202	0.0014	0.222	26427	25.6
19	204	0.0510	0.223	29229	28.3
19.5	206	0.1011	0.224	31794	30.7

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
20	208	0.1518	0.225	34916	33.8
20.5	210	0.2030	0.226	37735	36.5
21	212	0.2547	0.228	41204	39.8
21.5	214	0.3071	0.229	44327	42.9
22	216	0.3602	0.231	48295	46.7
22.5	218	0.4142	0.233	51848	50.1
23	220	0.4691	0.235	56043	54.2
23.5	222	0.5252	0.238	59741	57.8
24	224	0.5827	0.241	63919	61.8
24.5	226	0.6418	0.245	67711	65.5
25	228	0.7028	0.249	71911	69.5
25.5	230	0.7661	0.253	75723	73.2
26	232	0.8319	0.259	79720	77.1
26.5	234	0.9005	0.264	83205	80.5
27	236	0.9722	0.270	86653	83.8
27.5	238	1.0469	0.276	89547	86.6
28	240	1.1247	0.281	92205	89.2
28.5	242	1.2051	0.285	94566	91.4
29	244	1.2873	0.288	96600	93.4
29.5	246	1.3706	0.289	98200	95.0
30	248	1.4542	0.288	99519	96.2
30.5	250	1.5372	0.287	100472	97.2
31	252	1.6194	0.285	101268	97.9
31.5	254	1.7007	0.284	101873	98.5
32	256	1.7811	0.283	102329	99.0
32.5	258	1.8611	0.282	102641	99.3
33	260	1.9410	0.283	102895	99.5
33.5	262	2.0214	0.284	103077	99.7
34	264	2.1030	0.286	103186	99.8
34.5	266	2.1863	0.290	103261	99.9
35	268	2.2721	0.295	103313	99.9
35.5	270	2.3613	0.302	103347	99.9
36	272	2.4554	0.311	103377	100.0
36.5	274	2.5561	0.323	103394	100.0
37	276	2.6662	0.340	103401	100.0
37.5	278	2.7904	0.365	103406	100.0
38	280	2.9364	0.401	103411	100.0
38.5	282	3.1198	0.459	103413	100.0
39	284	3.3769	0.564	103414	100.0
39.5	286	3.8309	0.822	103414	100.0
40	288	4.7035	1.610	103414	100.0

TABLE C.3.2 Conversion 2004 NJ ASK Mathematics – Grade 3

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
0	119	-3.8535	1.803	2	0.0
0.5	121	-2.7099	0.955	7	0.0
1	124	-2.0996	0.650	17	0.0
1.5	126	-1.7656	0.517	26	0.0
2	129	-1.5386	0.440	52	0.1
2.5	131	-1.3670	0.390	80	0.1
3	133	-1.2290	0.354	133	0.1
3.5	136	-1.1131	0.327	195	0.2
4	138	-1.0129	0.306	295	0.3
4.5	140	-0.9244	0.289	430	0.4
5	143	-0.8447	0.275	627	0.6
5.5	145	-0.7721	0.263	861	0.8
6	148	-0.7051	0.254	1167	1.1
6.5	150	-0.6427	0.245	1571	1.5
7	152	-0.5841	0.238	2051	2.0
7.5	155	-0.5288	0.232	2613	2.5
8	157	-0.4763	0.226	3226	3.1
8.5	160	-0.4260	0.221	3916	3.8
9	162	-0.3779	0.217	4704	4.5
9.5	164	-0.3315	0.213	5587	5.4
10	167	-0.2866	0.210	6441	6.2
10.5	169	-0.2431	0.207	7418	7.2
11	171	-0.2008	0.204	8448	8.2
11.5	174	-0.1594	0.202	9595	9.3
12	176	-0.1190	0.200	10808	10.4
12.5	179	-0.0794	0.198	12031	11.6
13	181	-0.0404	0.196	13294	12.8
13.5	183	-0.0020	0.195	14753	14.2
14	186	0.0358	0.194	16245	15.7
14.5	188	0.0732	0.193	17712	17.1
15	190	0.1103	0.192	19452	18.8
15.5	193	0.1471	0.191	21075	20.4
16	195	0.1837	0.191	22878	22.1
16.5	198	0.2201	0.190	24260	23.4

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
17	200	0.2564	0.190	26412	25.5
17.5	202	0.2927	0.190	28454	27.5
18	205	0.3289	0.190	30550	29.5
18.5	207	0.3653	0.190	32680	31.6
19	210	0.4018	0.191	34994	33.8
19.5	212	0.4384	0.191	37370	36.1
20	214	0.4753	0.192	39757	38.4
20.5	217	0.5126	0.193	42233	40.8
21	219	0.5502	0.194	44782	43.2
21.5	221	0.5882	0.195	47352	45.7
22	224	0.6267	0.197	50117	48.4
22.5	226	0.6659	0.198	52828	51.0
23	229	0.7058	0.200	55738	53.8
23.5	231	0.7465	0.202	58621	56.6
24	233	0.7881	0.205	61730	59.6
24.5	236	0.8309	0.208	64567	62.3
25	238	0.8749	0.211	67827	65.5
25.5	240	0.9204	0.215	70621	68.2
26	243	0.9678	0.219	73932	71.4
26.5	245	1.0172	0.225	76701	74.1
27	248	1.0693	0.231	79934	77.2
27.5	250	1.1247	0.239	82585	79.7
28	252	1.1840	0.248	85758	82.8
28.5	255	1.2483	0.259	88176	85.1
29	257	1.3191	0.273	91157	88.0
29.5	260	1.3984	0.290	93174	90.0
30	262	1.4891	0.313	95822	92.5
30.5	264	1.5962	0.342	97379	94.0
31	267	1.7275	0.384	99488	96.1
31.5	269	1.8982	0.446	100536	97.1
32	271	2.1432	0.553	102204	98.7
32.5	274	2.5788	0.804	102666	99.1
33	276	3.4099	1.570	103559	100.0

TABLE C.4.1 Conversion 2004 NJ ASK Language Arts Literacy – Grade 4

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
0	106	-2.9922	1.504	14	0.0
0.5	109	-2.2304	0.767	15	0.0
1	112	-1.8402	0.517	69	0.1
1.5	114	-1.6308	0.407	76	0.1
2	117	-1.4915	0.343	170	0.2
2.5	121	-1.3885	0.300	177	0.2
3	124	-1.3071	0.271	315	0.3
3.5	127	-1.2397	0.249	343	0.3
4	131	-1.1817	0.232	498	0.5
4.5	134	-1.1307	0.219	545	0.5
5	137	-1.0846	0.209	719	0.7
5.5	140	-1.0423	0.201	781	0.8
6	143	-1.0028	0.195	1008	1.0
6.5	146	-0.9656	0.190	1111	1.1
7	148	-0.9300	0.186	1323	1.3
7.5	151	-0.8956	0.184	1472	1.4
8	153	-0.8621	0.182	1720	1.7
8.5	156	-0.8291	0.181	1921	1.9
9	158	-0.7965	0.180	2234	2.2
9.5	160	-0.7639	0.180	2495	2.4
10	162	-0.7311	0.181	2855	2.8
10.5	164	-0.6980	0.182	3171	3.1
11	166	-0.6643	0.184	3650	3.5
11.5	168	-0.6301	0.186	4061	3.9
12	170	-0.5951	0.188	4620	4.5
12.5	172	-0.5592	0.190	5147	5.0
13	174	-0.5225	0.192	5812	5.6
13.5	176	-0.4849	0.194	6480	6.2
14	178	-0.4465	0.196	7275	7.0
14.5	180	-0.4074	0.198	8092	7.8
15	182	-0.3676	0.200	9031	8.7
15.5	185	-0.3273	0.201	10086	9.7
16	187	-0.2865	0.202	11211	10.8
16.5	189	-0.2453	0.203	12403	11.9
17	191	-0.2039	0.203	13776	13.3
17.5	193	-0.1622	0.204	15282	14.7
18	195	-0.1204	0.204	16842	16.2
18.5	197	-0.0784	0.205	18546	17.9
19	200	-0.0362	0.205	20517	19.8
19.5	201	0.0061	0.206	22537	21.7
20	203	0.0487	0.206	24706	23.8
20.5	205	0.0914	0.207	27069	26.1
21	207	0.1344	0.207	29570	28.5
21.5	209	0.1778	0.208	32192	31.0

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
22	211	0.2215	0.209	34936	33.7
22.5	213	0.2658	0.211	37809	36.4
23	215	0.3106	0.212	40919	39.4
23.5	216	0.3560	0.214	44091	42.5
24	218	0.4022	0.215	47453	45.7
24.5	220	0.4491	0.217	50849	49.0
25	222	0.4970	0.219	54364	52.4
25.5	224	0.5458	0.222	57987	55.9
26	225	0.5957	0.224	61738	59.5
26.5	227	0.6466	0.227	65330	62.9
27	229	0.6987	0.229	68956	66.4
27.5	230	0.7519	0.231	72510	69.8
28	232	0.8062	0.234	75898	73.1
28.5	234	0.8616	0.236	79246	76.3
29	236	0.9179	0.238	82497	79.5
29.5	237	0.9751	0.240	85405	82.3
30	239	1.0330	0.241	88136	84.9
30.5	241	1.0915	0.242	90543	87.2
31	242	1.1505	0.243	92735	89.3
31.5	244	1.2099	0.244	94685	91.2
32	245	1.2696	0.244	96381	92.8
32.5	247	1.3295	0.245	97758	94.2
33	249	1.3897	0.245	98966	95.3
33.5	250	1.4503	0.246	99982	96.3
34	252	1.5114	0.247	100808	97.1
34.5	253	1.5731	0.249	101476	97.7
35	255	1.6358	0.251	102050	98.3
35.5	256	1.6996	0.254	102461	98.7
36	258	1.7651	0.257	102811	99.0
36.5	260	1.8325	0.262	103082	99.3
37	261	1.9025	0.267	103299	99.5
37.5	263	1.9758	0.274	103442	99.6
38	265	2.0533	0.282	103576	99.8
38.5	267	2.1361	0.293	103657	99.8
39	269	2.2261	0.306	103706	99.9
39.5	271	2.3253	0.324	103745	99.9
40	274	2.4374	0.346	103778	100.0
40.5	276	2.5677	0.377	103795	100.0
41	279	2.7259	0.420	103811	100.0
41.5	282	2.9299	0.487	103815	100.0
42	286	3.2222	0.604	103815	100.0
42.5	289	3.7454	0.883	103817	100.0
43	293	4.7461	1.713	103818	100.0

TABLE C.4.2 Conversion 2004 NJ ASK Mathematics – Grade 4

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
0	103	-4.4892	1.879	0	0.0
0.5	106	-3.1619	1.073	0	0.0
1	109	-2.3535	0.762	2	0.0
1.5	112	-1.8920	0.606	4	0.0
2	116	-1.5843	0.508	17	0.0
2.5	119	-1.3600	0.442	33	0.0
3	123	-1.1857	0.394	66	0.1
3.5	126	-1.0440	0.359	113	0.1
4	130	-0.9247	0.332	188	0.2
4.5	133	-0.8216	0.310	312	0.3
5	136	-0.7308	0.292	484	0.5
5.5	139	-0.6493	0.278	703	0.7
6	142	-0.5754	0.265	1002	1.0
6.5	144	-0.5076	0.255	1342	1.3
7	147	-0.4448	0.246	1768	1.7
7.5	150	-0.3862	0.238	2238	2.2
8	152	-0.3312	0.231	2794	2.7
8.5	155	-0.2793	0.224	3429	3.3
9	157	-0.2301	0.219	4133	4.0
9.5	159	-0.1833	0.213	4843	4.7
10	162	-0.1386	0.209	5687	5.5
10.5	164	-0.0957	0.205	6552	6.3
11	166	-0.0544	0.201	7504	7.2
11.5	168	-0.0147	0.197	8494	8.2
12	170	0.0236	0.194	9574	9.2
12.5	172	0.0608	0.191	10717	10.3
13	175	0.0967	0.188	11890	11.5
13.5	177	0.1317	0.185	13009	12.5
14	179	0.1657	0.183	14245	13.7
14.5	181	0.1988	0.180	15516	15.0
15	183	0.2311	0.178	16868	16.3
15.5	185	0.2627	0.176	18254	17.6
16	187	0.2936	0.174	19694	19.0
16.5	189	0.3238	0.173	21067	20.3
17	191	0.3535	0.171	22596	21.8
17.5	193	0.3826	0.169	24107	23.2
18	194	0.4111	0.168	25750	24.8
18.5	196	0.4392	0.167	27324	26.3
19	198	0.4669	0.165	28950	27.9
19.5	200	0.4941	0.164	30520	29.4
20	202	0.5210	0.163	32183	31.0
20.5	204	0.5476	0.162	33836	32.6
21	206	0.5738	0.161	35647	34.4
21.5	208	0.5998	0.160	37340	36.0

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
22	210	0.6255	0.160	39262	37.8
22.5	212	0.6510	0.159	40991	39.5
23	214	0.6763	0.158	42856	41.3
23.5	216	0.7014	0.158	44642	43.0
24	217	0.7265	0.158	46508	44.8
24.5	219	0.7514	0.157	48361	46.6
25	221	0.7763	0.157	50268	48.4
25.5	223	0.8011	0.157	52142	50.2
26	225	0.8260	0.157	54068	52.1
26.5	227	0.8508	0.157	55899	53.9
27	229	0.8758	0.158	57852	55.8
27.5	231	0.9009	0.158	59748	57.6
28	233	0.9262	0.159	61772	59.5
28.5	235	0.9517	0.160	63605	61.3
29	237	0.9774	0.161	65636	63.3
29.5	239	1.0035	0.162	67412	65.0
30	240	1.0300	0.163	69440	66.9
30.5	242	1.0570	0.164	71252	68.7
31	244	1.0844	0.166	73321	70.7
31.5	246	1.1125	0.168	75064	72.3
32	248	1.1413	0.170	77085	74.3
32.5	250	1.1709	0.173	78811	75.9
33	252	1.2015	0.176	80766	77.8
33.5	253	1.2331	0.179	82465	79.5
34	255	1.2660	0.183	84348	81.3
34.5	257	1.3002	0.187	85928	82.8
35	259	1.3361	0.191	87797	84.6
35.5	260	1.3739	0.197	89258	86.0
36	262	1.4139	0.202	90971	87.7
36.5	264	1.4564	0.209	92314	89.0
37	265	1.5019	0.217	93955	90.5
37.5	267	1.5510	0.226	95204	91.7
38	269	1.6046	0.236	96643	93.1
38.5	270	1.6636	0.249	97705	94.2
39	272	1.7293	0.264	98977	95.4
39.5	273	1.8038	0.282	99857	96.2
40	275	1.8900	0.305	100905	97.2
40.5	276	1.9924	0.335	101558	97.9
41	277	2.1189	0.377	102402	98.7
41.5	279	2.2841	0.439	102816	99.1
42	280	2.5226	0.546	103359	99.6
42.5	282	2.9493	0.797	103524	99.8
43	285	3.7686	1.561	103770	100.0

REFERENCES

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (AERA, APA, NCME) (1999) *Standards for Educational and Psychological Testing*. Washington, DC: Author.
- Camilli, G. & Shepard, L.A. (1994). *Methods for Identifying Biased Test Items*. Thousand Oaks, CA: Sage.
- Cizek, G.J. (1995). *Standard-Setting as Psychometric Due Process*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.
- DeMauro, G.E. (1995). *Construct Validation of Minimum Competence*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.
- Directory of Test Specifications and Sample Items for the Elementary School Proficiency Assessment (ESPA) in Mathematics*. Developed by the New Jersey Department of Education, April 1998.
- Directory of Test Specifications and Sample Items for the Elementary School Proficiency Assessment (ESPA) in Science*. Developed by the New Jersey Department of Education, April 1998.
- Directory of Test Specifications and Sample Items for the Elementary School Proficiency Assessment (ESPA), Grade Eight Proficiency Assessment (GEPA), and High School Proficiency Assessment (HSPA) in Language Arts Literacy*. Developed by the New Jersey Department of Education, February 1998.
- Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and the High School Proficiency Assessment (HSPA) in Mathematics*. Developed for the New Jersey Department of Education, February 1998.
- Elementary School Proficiency Assessment (ESPA) Standard Setting Summary. 1999 Language Arts Literacy, Mathematics and Science*. Developed for the New Jersey Department of Education by National Computer Systems, August 1999.
- Grade Eight Proficiency Assessment (GEPA) Standard Setting Study Report of Activities. 1999 Language Arts Literacy and Mathematics, and 2000 Science*. Developed for the New Jersey Department of Education by National Computer Systems, November 2000.
- Hambleton, R., Swaminathan, H., Rogers, H. (1991). *Fundamentals of Item Response Theory*. Newbury Park: Sage Publications, Inc.
- Holland, P.W. & Thayer, D.T. (1986). *Differential Item Functioning and the Mantel-Haenszel Procedure* (Technical Rep. No. 86-69). Princeton, NJ: Educational Testing Service.
- Jaeger, R.M., & Mills, C.N. (1998). An integrated judgment procedure for setting standards on complex large-scale assessments. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

- Kolen, M.J. & Brennan, R.L. (1995). *Test Equating*. New York: Springer-Verlag.
- Linn, R. L. (Ed.). (1989). *Educational Measurement* (3rd ed.). New York: American Council on Education.
- Livingston, S.A. & Zieky, M.J. (1982) *Passing Scores*. Princeton, NJ: Educational Testing Service.
- Livingston, S. A., and Lewis, C. (1995). Estimating the consistency and accuracy of classification based on test scores. Journal on Educational Measurement, 32, 179-197.
- Mehrens, W.A. & Lehmann, I.J. (1991) *Measurement and Evaluation in Education and Psychology* (4th ed.). New York: Holt, Rinehart and Winston.
- Millman, J., & Greene, J. (1989). The specification and development of test of achievement and ability. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 335-366). New York: Macmillan.
- SAS Institute (1985). *User's Guide: Statistics* (vol. I and II). Cary, NC: Author.
- Van der Linden, W.J. & Hambleton, R.K. (1997). *Handbook of Modern Item Response Theory*. New York: Springer-Verlag.
- Webb, M.W. & Miller, E.R. (1995). *Standard Setting On Constructed-Response Items*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P.W. Holland and H. Wainer (Eds.), *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum.