

Grade 4  
New Jersey Assessment of Skills  
and Knowledge

**TECHNICAL REPORT**

**MAY 2003**

PTM 1505.25  
Copyright © 2005 by New Jersey Department of Education  
All rights reserved

## TABLE OF CONTENTS

<b>PART 1: INTRODUCTION .....</b>	<b>1</b>
1.1 Description of the New Jersey Assessment of Skills and Knowledge (NJ ASK) .....	1
1.2 State-Level Results .....	1
1.3 NJ ASK Organizational Support.....	2
<b>PART 2: TEST DEVELOPMENT .....</b>	<b>3</b>
2.1 Test Specifications .....	3
Language Arts Literacy .....	4
Mathematics.....	4
2.2 Development of Test Items .....	5
2.3 Item Review Process.....	7
2.4 Item Use.....	9
2.5 Test Forms Assembly .....	9
2.6 Quality Control for Test Construction.....	10
<b>PART 3: TEST ADMINISTRATION .....</b>	<b>10</b>
3.1 Participation .....	11
3.2 Test Security Procedures .....	11
3.3 Test Administration Procedures.....	12
3.4 Test Accommodations .....	12
<b>PART 4: SCORING.....</b>	<b>14</b>
4.1 Multiple Choice Items .....	14
4.2 Open Ended Items.....	14
Scorer Selection .....	14
Range Finding.....	14
Scorer Training .....	15
Scoring Procedures .....	15
4.3 Quality Control Procedures in Data Preparation .....	16
<b>PART 5: STANDARD SETTING .....</b>	<b>16</b>
<b>PART 6: ITEM-LEVEL STATISTICS.....</b>	<b>17</b>
6.1 Classical Item Statistics .....	17
6.2 Speededness .....	22
6.3 Intercorrelations .....	22
6.4 Item Bias Statistics.....	26
<b>PART 7: SCALING AND EQUATING .....</b>	<b>27</b>
7.1 Scale Scores .....	27
7.2 Equating Language Arts Literacy .....	28
7.3 Equating Mathematics .....	31
<b>PART 8: TEST STATISTICS .....</b>	<b>33</b>
8.1 Summary Statistics .....	33
8.2 Classical Reliability Estimates of the Test Scores.....	38
8.3 Reliability of Performance Classification.....	41
8.4 Conditional Estimate of Error at Each Cut-Score.....	43
8.5 Rater Reliability .....	43

<b>Part 9: Validity .....</b>	<b>46</b>
Content and Curricular Validity .....	46
<b>Part 10: Reporting .....</b>	<b>46</b>
10.1 Cycle I Reports .....	46
10.2 Cycle II Reports .....	51
10.3 Interpreting Reports .....	54
10.4 Quality Control in Reporting .....	54
<b>APPENDIX A: State Summary .....</b>	<b>55</b>
<b>APPENDIX B: District Factor Groups.....</b>	<b>68</b>
<b>APPENDIX C: Raw Score – Scale Score Conversions with Frequencies .....</b>	<b>70</b>
<b>REFERENCES.....</b>	<b>72</b>

## Tables and Figures

TABLE 1.2.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Total Student Group Tested.....	2
TABLE 2.1.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Total Points Possible by Content Area.....	5
Figure 2.3.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Item Approval Before Field Test.....	7
Figure 2.3.2 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Item Approval Before Operational Base Test.....	8
TABLE 5.1.1 ESPA Grade 4 Standard-Setting Results Recommended Cut Scores for Language Arts Literacy and Mathematics .....	17
TABLE 6.1.1 2003 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 .....	19
TABLE 6.1.2 2003 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.....	20
TABLE 6.1.3 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Frequency Distributions of Item Difficulty Values and Biserial Discrimination Indices by Content Cluster: Mathematics .....	21
TABLE 6.2.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Percentage of Students Omitting the Last Items of Each Test Part.....	23
TABLE 6.3.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Intercorrelations Among Major Content Clusters and Item Types.....	24
TABLE 6.3.2 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Intercorrelations Among Content Areas and Clusters .....	25
TABLE 6.4.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) DIF Categories .....	27
TABLE 7.1.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Scaling Parameters for Base Forms .....	28
Figure 7.2.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Language Arts Literacy Backwards and Forward Equating Designs.....	29
TABLE 7.2.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Language Arts Literacy Item Parameters .....	30
TABLE 7.2.2 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Language Arts Literacy Fixed OE Item Step Parameters.....	30
TABLE 7.3.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Mathematics Item Parameters .....	32
TABLE 7.3.2 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Mathematics Fixed OE Item Step Parameters.....	33

TABLE 8.1.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students' Raw Scores by Test Section.....	34
TABLE 8.1.2 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Means and Standard Deviations of Students' Raw Scores and Percent Correct by Content Area.....	35
TABLE 8.1.3 2003 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 .....	36
TABLE 8.1.4 2003 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 Cluster by Content Area.....	37
TABLE 8.2.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability Estimates and Standard Errors of Measurement (SEM) for Content Areas and Clusters ....	39
TABLE 8.2.2 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability Estimates and Standard Errors of Measurement (SEM) for Dichotomously Scored Items Within Content Clusters.....	40
TABLE 8.3.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Reliability of Classification and Decision Accuracy .....	42
TABLE 8.4.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Conditional Estimate of Error at Each Cut-Score.....	43
TABLE 8.5.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Consistency Between Raters Scoring Writing Tasks and Open-Ended Items .....	45
Figure 10.1.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Individual Student Report (ISR) – Front.....	48
Figure 10.1.2 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) Individual Student Report (ISR) – Back .....	49
Figure 10.2.1 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) School Report - Performance by Demographic Groups.....	52
Figure 10.2.2 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) School Report - Cluster Score Means .....	53
TABLE A.1 NJ ASK Statewide Performance by Demographic Groups Spring 2003 Language Arts Literacy .....	57
TABLE A.2 NJ ASK Statewide Performance by Demographic Groups Spring 2003 Mathematics.....	58
TABLE A.3 New Jersey Statewide Testing System Spring 2003 New Jersey Assessment of Skills and Knowledge By District Factor Group Language Arts Literacy .....	60
TABLE A.4 New Jersey Statewide Testing System Spring 2003 New Jersey Assessment of Skills and Knowledge By District Factor Group Mathematics Section .....	63
TABLE A.5 New Jersey Statewide Testing System Spring 2003 New Jersey Assessment of Skills and Knowledge Language Arts Literacy and Mathematics for The Special Needs Districts as Compared to All Other Districts .....	66

## **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 [May 2003](#). 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, 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 2003 New Jersey Assessment of Knowledge and Skills (NJ ASK) was administered operationally to students in grade four. Grade 3 was administered as a field-test to grade three students in 2003. Thus, grade 3 results will not be presented here. The 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 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 May 19 and May 30, 2003. The Language Arts Literacy test was administered to 106,286 total students, and Mathematics was administered to 106,134 total students.

### **1.2 State-Level Results**

This section includes a table summarizing statewide test results for the 2003 administration of the NJ ASK. Table 1.2.1 shows 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. The “number of students tested” is based on all students who received a test booklet, excluding those who were voided or IEP exempt with no scale scores. (IEP stands for Individual Education Program.)

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

Following are two state-level highlights for all students.

- Of the 106,286 grade 4 students with valid scale scores in Language Arts Literacy in Spring 2003, 22.4% scored in Partially Proficient; 73.8% scored in Proficient and 3.8% scored in Advanced Proficient (Table 1.2.1).
- Of the 106,134 grade 4 students with valid scale scores in Mathematics in Spring 2003, 32.0% scored in Partially Proficient; 42.8% scored in Proficient and 25.2% scored in Advanced Proficient (Table 1.2.1).

**TABLE 1.2.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Total Student Group Tested**

TEST SECTION	NUMBER <sup>a</sup> TESTED	PROFICIENCY LEVELS						SCALE SCORE MEAN
		PARTIALLY PROFICIENT (100-199)		PROFICIENT (200-249)		ADVANCED PROFICIENT (250-300)		
		No	%	No	%	No	%	
LANGUAGE ARTS LITERACY 2003	106,286	23,802	22.4	78,418	73.8	4,066	3.8	214.6
MATHEMATICS 2003	106,134	33,985	32.0	45,414	42.8	26,735	25.2	217.3

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID, AND IEP 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 The Grow Network and Riverside Publishing Company. The major ETS activities include program management; test development; publication development; 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, 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 Grow Network is responsible for producing, printing and shipping reports of test results for 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 1999 through 2002 to provide an early indication of student progress toward achieving the knowledge and skills identified in the Core Curriculum Content Standards (CCCS). The ESPA was replaced in spring 2003 with the New Jersey Assessment of Skills and Knowledge (NJ ASK), a comprehensive, multi-grade assessment program. 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 items 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 knowledge and problem-solving ability. Each test question on the Mathematics assessment measures one content cluster and may contribute to the process cluster.

Five open-ended items appear at Grade 4. For an in-depth description of the NJ ASK Mathematics Test Specifications visit the NJ Department of Education website at:

<http://www.njpep.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 [May 2003](#).

**TABLE 2.1.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Total Points Possible by Content Area**

<b>Language Arts Literacy</b>	
<b>Total</b>	<b>43 points</b>
Reading	23 points
Writing	20 points
Writing/Picture	10 points
Writing/Poem	10 points
<hr style="border-top: 1px dashed black;"/>	
Working with Text	8 points
Analyzing Text	15 points
<b>Mathematics</b>	
<b>Total</b>	<b>42 points</b>
C4.1 - Number Sense & Numerical Operations	13 points
C4.2 - Geometry & Measurement	10 points
C4.3 - Patterns & Algebra	9 points
C4.4 - Data Analysis, Probability & Discrete Math	10 points
<hr style="border-top: 1px dashed black;"/>	
Knowledge	42 points
Problem Solving	32 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 NJ ASK usually 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.

No grade 4 NJ ASK items were field-tested in 2003. 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 of the New Jersey Department of Education (NJDOE), Office of Evaluation and Assessment.

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**

**2003 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**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Item Approval Before Operational Base Test**

Sensitivity	Content
<b>*Comments</b>	<b>*Comments</b>
Sensitivity Issue                      Yes              No If Yes, identify category and explain*	Appropriate Difficulty                      Yes              No P-Value = 0.65 Biserial = 0.42
Mantel-Haenszel Category C  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

## 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. 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 form 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). 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. 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 ensure that 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 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) included testing sections in Language Arts Literacy and Mathematics. 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.

Language Arts Literacy and Mathematics field-test items are usually embedded within the sections of the regular test. 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 are used to 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 her or him from taking the NJ ASK must participate in the IEP. 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 2003, 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. Preference is given to individuals with degrees and backgrounds related to language arts, mathematics and science, and experience in performance scoring. Scoring leaders are chosen based on subject area expertise, along with strong organizational abilities and communication skills. Scoring leaders must demonstrate the ability to assist Content Scoring Leaders 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. 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 scoring leaders to manage scoring performance and to identify changes or trends in the scorer's performance. Scoring leaders work with scorers assigned to their shift to ensure scoring quality.

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.

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.

## **PART 5: STANDARD SETTING**

Standard setting is the process used to establish cut scores on a test score scale that allows test score users to distinguish performance among various levels. On the NJ ASK, students score in one of three performance levels: Partially Proficient, Proficient, and Advanced Proficient. The term *base form* is used to describe the test form on which the raw-to-scale score conversion was originally specified. Standards were set on the base form of the Mathematics Elementary School Proficiency Assessment (ESPA) in 1999. Standards were set on the base form of the Language Arts Literacy ESPA in 2001.

Through equating and scaling procedures, the standards established on the base forms have been carried forward so that the amount of ability needed to score in Proficient and Advanced Proficient performance levels has remained the same each year despite small changes in the difficulty of the test forms. For more information on the standards-setting studies for the ESPA, please see the document *Standard-Summary Procedures for New Jersey Statewide Assessments*.

Table 5.1.1 shows the final cut scores that were approved by the State Board of Education. In addition, the table shows the percentage of grade 4 students.

**TABLE 5.1.1**

**ESPA Grade 4 Standard-Setting Results  
Recommended Cut Scores for Language Arts Literacy and Mathematics**

	Points	Cut score for Proficient	Cut Score for Advanced Proficient	% Partially Proficient	% Proficient	% Advanced Proficient
<b>Math - 1999</b>	0-43	23.0	35.0	39.5%	44.3%	16.2%
<b>LAL - 2001</b>	0-43	21.5	33.0	21.1%	69.9%	9.0%

**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 0.30 to 0.95.
- Item discrimination (“r-biserial)<sup>1</sup>:  
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

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

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

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

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, if students have an adequate amount of testing time, 95% of students would be expected to 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, it may indicate that students had insufficient time to complete all items. Alternatively, if the omit percentage is greater than 5% for a single item, it could be an indication of an item/test layout problem. For example, students might accidentally skip an item that follows a lengthy stem.

In Table 6.1.1, summary statistics are given that describe the difficulty and discrimination of the items comprising each cluster. For dichotomously scored items, means and standard deviations of proportion-correct values (p-values) and point 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 point-biserial correlation, students' total test scores were expressed in terms of the raw score metric.

Frequency distributions of the May 2003 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 6.1.2 and 6.1.3, respectively. The top section of each table shows the distribution of item difficulty values; the bottom section shows the distribution of point-biserial indices.

**TABLE 6.1.1**

**2003 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**

NJ ASK Test Section/Cluster	Dichotomous			Open-Ended		
	Item Difficulty		Item Discrimination	Item Difficulty		Item Discrimination
	Mean	S.D.	Mean	Mean	S.D.	Mean
<b>Language Arts Literacy</b>	0.62	0.10	0.42	0.51	0.11	0.78
Reading	0.62	0.10	0.42	0.38	0.07	0.71
Writing	--	--	--	0.59	0.02	0.82
Writing/Picture	--	--	--	0.60	--	0.82
Writing/Poem	--	--	--	0.57	--	0.81
Working with Text	0.64	0.10	0.41	--	--	--
Analyzing Text	0.56	0.10	0.44	0.38	0.07	0.71
<b>Mathematics</b>	0.70	0.12	0.42	0.46	0.18	0.71
Number Sense & Numerical Operations	0.75	0.10	0.40	0.57	0.11	0.72
Geometry & Measurement	0.68	0.13	0.42	0.18	--	0.66
Patterns & Algebra	0.74	0.15	0.40	0.45	--	0.75
Data Analysis, Probability & Discrete Math	0.63	0.09	0.45	0.55	--	0.69
Knowledge	0.70	0.12	0.42	0.46	0.18	0.71
Problem Solving	0.69	0.13	0.43	0.46	0.18	0.71

**TABLE 6.1.2**

**2003 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**

<b>Item Statistics</b>	<b>Working With Text</b>	<b>Analyzing Text</b>	<b>Total</b>
<b>ITEM DIFFICULTY: P-VALUES</b>			
.800 - .899	0	0	0
.700 - .799	3	0	3
.600 - .699	3	1	4
.500 - .599	1	1	2
<.500	1	1	2
<b>MEAN P-VALUE</b>	0.64	0.56	0.62
<b>MEDIAN P-VALUE</b>	0.66	0.56	0.62
<b>ITEM DISCRIMINATION: POINT-BISERIAL CORRELATIONS</b>			
.50+	0	1	1
.40 - .49	6	2	8
.30 - .39	2	0	2
<b>MEAN POINT-BISERIAL</b>	0.41	0.44	0.42
<b>MEDIAN POINT-BISERIAL</b>	0.42	0.43	0.42
<b>TOTAL NUMBER OF ITEMS</b>	8	3	11

**TABLE 6.1.3**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Frequency Distributions of Item Difficulty Values and Biserial Discrimination Indices  
by Content Cluster: Mathematics**

<b>Item Statistics</b>	<b>Number Sense &amp; Numerical Operations</b>	<b>Geometry &amp; Measurement</b>	<b>Patterns &amp; Algebra</b>	<b>Data Analysis, Probability &amp; Discrete Math</b>	<b>Knowledge</b>	<b>Problem Solving</b>	<b>Total Tests</b>
<b>ITEM DIFFICULTY: P-VALUES</b>							
.900 - .999	2	0	1	0	3	1	3
.800 - .899	2	2	1	0	5	3	5
.700 - .799	6	0	2	2	10	4	10
.600 - .699	0	3	1	2	6	4	6
.500 - .599	1	2	1	2	6	4	6
<.500	0	0	0	1	1	1	1
<b>MEAN P-VALUE</b>	0.80	0.68	0.74	0.63	0.68	0.69	0.68
<b>MEDIAN P-VALUE</b>	0.76	0.66	0.74	0.65	0.71	0.66	0.71
<b>ITEM DISCRIMINATION: POINT-BISERIAL CORRELATIONS</b>							
.50 - .59	1	1	1	2	5	5	5
.40 - .49	6	3	2	3	14	6	14
.30 - .39	3	3	3	2	11	6	11
.20 - .29	1	0	0	0	1	0	1
<b>MEAN POINT-BISERIAL</b>	0.40	0.42	0.40	0.45	0.42	0.43	0.42
<b>MEDIAN POINT-BISERIAL</b>	0.43	0.40	0.39	0.48	0.42	0.45	0.42
<b>Total Number of Items</b>	11	7	6	7	31	17	31

## 6.2 Speededness

The NJ ASK is intended to provide sufficient time for all students to respond to almost all of the questions. Table 6.2.1 presents data concerning the extent to which this intent was met. Table 6.2.1 shows that the percent of students omitting the Reading multiple-choice items was very small while the percent of students omitting the open-ended items varied from 2.0% to 6.2%.

Table 6.2.1 also shows the percentage of students omitting each of the last two Mathematics multiple-choice items in each part and all Mathematics open-ended items. The percent of students omitting the Mathematics multiple-choice items ranged from 0.5% to 4.3%. The percent of students omitting the Mathematics open-ended items ranged from 1.2% to 4.0%.

## 6.3 Intercorrelations

Pearson product-moment correlation between student scores on the Language Arts Literacy and Mathematics content areas is .71. Table 6.3.1 also shows the correlations between students' scores in the major content clusters and item types. Table 6.3.1 shows 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. Clusters with more items that make up a higher proportion of the content area score increase the cluster-area correlation. For example, the correlation between Mathematics Total and Mathematics Multiple-Choice in Table 6.3.1 is quite high at .96 because 27 Mathematics Multiple-Choice points are part of the total Mathematics Total 42 points.

In addition, correlations are partially a function of the number of items in the measures being correlated. Therefore, the number of items in the content areas and clusters being correlated must be considered when their correlations are evaluated.

**TABLE 6.2.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Percentage of Students Omitting the Last Items of Each Test Part**

Test Section	Multiple – Choice		Open - Ended	
	Item Number	Percentage Omitting	Item Number	Percentage Omitting
<b>Reading</b>				
<u>First Part</u>	Item 4	0.6%	Item 6	2.0%
	Item 5	0.7%	Item 7	6.2%
<u>Second Part</u>	Item 5	1.0%		
	Item 6	1.0%	Item 7	2.6%
<b>Mathematics</b>				
Day 1 <u>First Part</u>	Item 3	2.2%		
	Item 4	4.3%		
<u>Second Part</u>	Item 7	0.5%		
	Item 8	0.9%		
<u>Third Part</u>	Item 19	2.4%	Item 21	3.8%
	Item 20	3.1%		
<u>Fourth Part</u>	Item 26	1.0%	Item 28	1.8%
	Item 27	1.1%	Item 29	4.0%
Day 2 <u>Fifth Part</u>	Item 33	0.6%	Item 36	1.2%
	Item 34	1.1%	Item 37	3.6%

Item 35 in Mathematics was “Do not score”

**TABLE 6.3.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Intercorrelations Among Major Content Clusters and Item Types**

Major Content Clusters and Item Types	Major Content Clusters and Item Types							
	Language Arts Literacy (LAL)					Mathematics (MAT)		
	LAT	R	R MC	R OE	W	MAT	M MC	M OE
<b>LAL Language Arts Literacy (43)</b>								
R Reading (23)	.94							
R MC Reading Multiple-Choice (11)	.83	.92						
R OE Reading Open-ended (12)	.86	.89	.63					
W Writing (20)	.89	.67	.56	.66				
<b>MAT Mathematics (42)</b>	.71	.71	.65	.63	.57			
M MC Mathematics Multiple-Choice (27)	.68	.68	.63	.60	.55	.96		
M OE Mathematics Open-ended (15)	.65	.66	.59	.59	.53	.92	.77	

Number in Parentheses is the number of points.  
Language Arts Literacy N=106,165; Mathematics N=106,016.

**TABLE 6.3.2**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Intercorrelations Among Content Areas and Clusters**

Test Section/Cluster	Test Section/Cluster													
	LAL Language Arts Literacy							MAT Mathematics						
	LAL	L1	L2	L3	L4	L5	L6	MAT	M1	M2	M3	M4	M5	M6
<b>LAL Language Arts Literacy (43)</b>														
L1 Reading (23)	.94													
L2 Writing (20)	.89	.67												
L3 Writing / Picture (10)	.81	.63	.89											
L4 Writing / Poem (10)	.80	.59	.91	.63										
L5 Working with Text (8)	.78	.86	.53	.50	.46									
L6 Analyzing Text (15)	.90	.94	.67	.62	.59	.64								
<b>MAT Mathematics (42)</b>	.71	.71	.57	.54	.49	.60	.68							
M1 Number Sense and Numerical Operations(13)	.64	.63	.53	.50	.46	.53	.60	.90						
M2 Geometry and Measurement (10)	.58	.59	.46	.44	.40	.50	.56	.83	.66					
M3 Data Analysis, Probability and Discrete Math (9)	.64	.65	.50	.48	.43	.55	.61	.87	.70	.66				
M4 Patterns and Algebra (10)	.59	.59	.47	.44	.41	.50	.56	.85	.69	.62	.65			
M5 Knowledge (42)	.71	.71	.57	.54	.49	.60	.68	1.00	.90	.83	.87	.85		
M6 Problem Solving (32)	.70	.70	.56	.53	.48	.59	.67	.99	.88	.83	.87	.84	.99	

Number in Parentheses is the number of points.  
Language Arts Literacy N=106,165; Mathematics N=106,016.

## 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**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
DIF Categories**

<b>DIF Category</b>	<b>Definition</b>
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 2003 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 2003 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. Table 7.1.1 shows the scaling parameters for each test.

**TABLE 7.1.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Scaling Parameters for Base Forms**

Grade	Subject	Base Year	Points	Slope	Intercept
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**

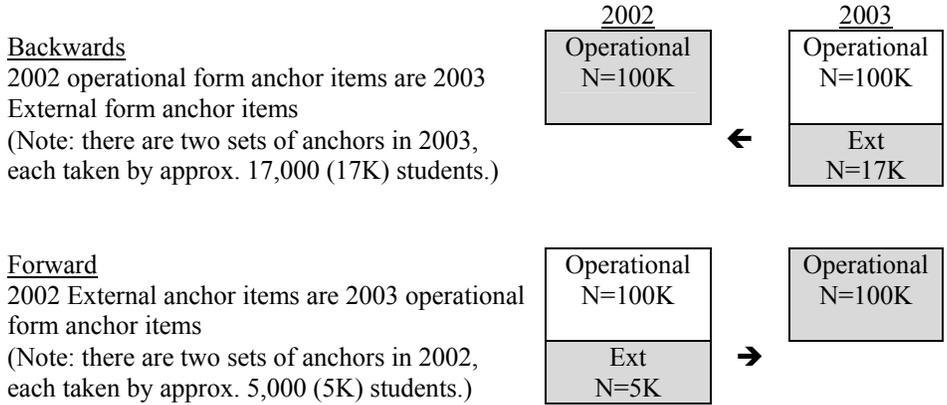
Scores on the 2003 NJ ASK grade 4 LAL form were equated back to scores on the 2001 LAL base form via 2002 anchored Rasch difficulty parameters and using IRT true score equating procedures. The grade 4 base year LAL raw score scale ranged from 0-43.0. The base year raw cut score for Proficient was 21.5 (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., 2002) and new (e.g., 2003) 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., 2002) and are in external sets on the new form (e.g., 2003). The Forward equating items were “pre-tested” as external sets on the old form (2002) and appear in the operational form on the new form (2003).

**Figure 7.2.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Language Arts Literacy Backwards and Forward Equating Designs**



Performance on the Backward equating anchor items in 2003 indicate students in 2003 were slightly less able than in 2002, and the 2003 form was slightly more difficult than the 2002 form. After comparing the results of these two equating approaches, the recommended raw-score to scale-score conversion for the 2003 NJ ASK LAL test resulted from the Backwards approach. The recommended raw score cut points in 2003 for LAL were 18.0 and 35.0 for Proficient and Advanced Proficient categories, respectively. Details about the methods and results are described in the 2003 LAL Equating Report. Table 7.2.1 shows the Rasch difficulty parameters (“Measure”), and item fit statistics from WINSTEPS for the Backwards 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 2003 may be found in Appendix B. To create a Braille form a committee reviewed the 2003 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**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Language Arts Literacy Item Parameters**

Item No.	Measure	Anchor	Error	IN FIT		OUT FIT		Score Corr.
				MNSQ	ZSTD	MNSQ	ZSTD	
1	0.0995	Anchor	0.0014	0.91	-9.9	0.92	-9.9	0.81
2	0.4912	Anchor	0.0035	1.18	9.9	1.32	9.9	0.30
3	0.1620	Anchor	0.0035	0.97	-9.9	0.97	-6.2	0.48
4	0.4711	Anchor	0.0035	1.06	9.9	1.13	9.9	0.41
5	0.0776	Anchor	0.0036	1.03	9.9	1.06	9.9	0.43
6	-0.1221	Anchor	0.0037	1.03	7.2	1.06	8.6	0.42
7	0.8203	Anchor	0.0023	0.72	-9.9	0.72	-9.9	0.76
8	1.2274	Anchor	0.0022	0.93	-9.9	0.91	-9.9	0.66
9	0.2524	Anchor	0.0013	1.04	8.4	1.06	9.9	0.79
10	-0.2343	Anchor	0.0039	1.00	-0.3	1.01	1.9	0.43
11	-0.2975	Anchor	0.0039	1.03	7.0	1.05	6.3	0.40
12	0.1246	Anchor	0.0035	1.08	9.9	1.12	9.9	0.39
13	-0.1705	Anchor	0.0038	1.03	9.1	1.10	9.9	0.40
14	-0.0387	Anchor	0.0037	0.94	-9.9	0.9	-9.9	0.49
15	0.2338	Anchor	0.0035	1.03	9.9	1.08	9.9	0.43
16	0.7613	Anchor	0.0020	1.01	2.9	1.01	2.3	0.67

**TABLE 7.2.2**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Language Arts Literacy Fixed OE Item Step Parameters**

Item	Category	Step	Item	Category	Step	Item	Category	Step
7	0	0	8	0	0	16	0	0
7	1	-0.79	8	1	-0.72	16	1	-0.31
7	2	-1.90	8	2	-1.51	16	2	-1.69
7	3	-0.62	8	3	-0.65	16	3	-0.36
7	4	-1.01	8	4	-0.45	16	4	-0.31
7	5	0.64	8	5	0.33	16	5	0.12
7	6	0.49	8	6	0.45	16	6	0.14
7	7	1.44	8	7	1.02	16	7	1.06
7	8	1.75	8	8	1.53	16	8	1.35

### 7.3 Equating Mathematics

Scores on the 2003 NJ ASK grade 4 Mathematics form were equated back to scores on the 1999 Mathematics base form via 2001 and 2002 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 1999.

The data collection design for the NJ ASK Mathematics test is also Common-Item Nonequivalent Groups design. The 2003 Mathematics test used both internal and external 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 2003, 23 items from 2001 and 2002 forms were used to link back to 1999. Eleven anchor items were embedded in the new form and 12 were located in external variable sections (Forms A and B). One item was flagged by NJDOE as "Do Not Score." As a result, scores on the 2003 NJ ASK grade 4 Mathematics test ranged from 0-42.

Based on the performance on the anchor items, the 2003 students appear to be about the same in ability as the 2002 students and the 2003 form was similar in difficulty to the 2002 form. The recommended raw-score (and scale-score) cut points for the 2003 Mathematics NJ ASK based on the equating results were 22.0 (200) and 33.0 (250) for Proficient and Advanced Proficient categories, respectively. Details about the methods and results are described in the 2003 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 2003 Mathematics test items. No items were dropped from the 2003 NJ ASK Mathematics operational form to create the Braille form. A separate raw-to-scale score conversion table was not needed for the Mathematics Braille form in 2003.

**TABLE 7.3.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Mathematics Item Parameters**

Item No.	Measure	Anchor	Error	IN FIT		OUT FIT		Score Corr.
				MNSQ	ZSTD	MNSQ	ZSTD	
1	-2.309	Free	0.0142	0.96	-3.6	0.84	-9.9	0.22
2	-0.734	Free	0.0081	0.89	-9.9	0.79	-9.9	0.43
3	-0.327	Free	0.0073	0.90	-9.9	0.85	-9.9	0.42
4	-0.604	Free	0.0078	0.88	-9.9	0.78	-9.9	0.45
5	-1.585	Free	0.0106	0.92	-9.9	0.73	-9.9	0.35
6	-0.948	Free	0.0086	0.94	-9.9	0.88	-9.9	0.32
7	-0.006	Anchor	0.0069	0.85	-9.9	0.83	-9.9	0.41
8	-0.221	Free	0.0072	0.90	-9.9	0.86	-9.9	0.42
9	0.223	Free	0.0038	1.19	9.9	1.28	9.9	0.30
10	0.383	Free	0.0037	1.11	9.9	1.16	9.9	0.37
11	0.194	Free	0.0039	1.09	9.9	1.15	9.9	0.37
12	0.245	Anchor	0.0038	1.15	9.9	1.18	9.9	0.38
13	0.027	Free	0.0041	0.92	-9.9	0.87	-9.9	0.47
14	0.549	Anchor	0.0036	1.03	9.9	1.07	9.9	0.44
15	0.811	Free	0.0036	0.91	-9.9	0.90	-9.9	0.52
16	0.661	Free	0.0036	0.99	-5.0	0.98	-4.3	0.47
17	-0.049	Anchor	0.0042	1.05	9.9	1.00	0.5	0.40
18	-0.184	Anchor	0.0045	0.98	-4.0	0.96	-3.9	0.39
19	0.318	Anchor	0.0037	0.94	-9.9	0.89	-9.9	0.50
20	0.325	Free	0.0037	0.97	-7.8	0.99	-1.2	0.46
21	0.704	Anchor	0.0020	1.00	-0.6	0.97	-6.3	0.71
22	0.332	Free	0.0037	1.12	9.9	1.17	9.9	0.36
23	-0.010	Anchor	0.0041	0.95	-9.9	0.99	-0.8	0.46
24	-0.550	Free	0.0054	1.02	3.1	1.03	2.0	0.31
25	0.575	Free	0.0036	1.08	9.9	1.11	9.9	0.40
26	0.380	Free	0.0037	0.95	-9.9	0.92	-9.9	0.49
27	0.762	Free	0.0036	0.93	-9.9	0.93	-9.9	0.51
28	0.833	Free	0.0018	1.09	9.9	1.12	9.9	0.67
29	0.656	Free	0.0019	1.15	9.9	1.18	9.9	0.64
30	-0.577	Free	0.0055	1.00	0.0	1.25	9.9	0.30
31	0.490	Anchor	0.0036	1.04	9.9	1.05	9.1	0.44
32	0.637	Anchor	0.0036	1.08	9.9	1.15	9.9	0.39
33	-0.618	Anchor	0.0057	0.92	-9.9	0.75	-9.9	0.39
34	0.183	Free	0.0039	1.13	9.9	1.14	9.9	0.34
35	0.379	Free	0.0020	1.12	9.9	1.23	9.9	0.63
36	1.579	Free	0.0026	1.11	9.9	1.12	9.9	0.55

**TABLE 7.3.2**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Mathematics Fixed OE Item Step Parameters**

Item	Category	Step	Item	Category	Step	Item	Category	Step
21	0	0	28	0	0	29	0	0
21	1	1.12	28	1	0.70	29	1	1.48
21	2	-1.61	28	2	0.19	29	2	-1.62
21	3	0.96	28	3	0.89	29	3	1.17
21	4	-1.49	28	4	-1.18	29	4	-1.38
21	5	1.55	28	5	0.90	29	5	2.16
21	6	-0.54	28	6	-1.50	29	6	-1.80
Item	Category	Step	Item	Category	Step			
35	0	0	36	0	0			
35	1	1.41	36	1	0.90			
35	2	-2.01	36	2	-1.70			
35	3	1.24	36	3	2.91			
35	4	-0.70	36	4	-1.20			
35	5	1.27	36	5	-0.09			
35	6	-1.21	36	6	-0.82			

**PART 8: TEST STATISTICS**

**8.1 Summary Statistics**

Means and standard deviations of students' raw scores on each content area are given in Table 8.1.1 for the May 2003 test. These data are based on the total Grade 4 student population described in Part 1. Table 8.1.1 shows that students' mean raw scores were 23.1 of 43 points for Language Arts Literacy, and 25.8 of 42 points for Mathematics. The table also shows that the standard deviations of the raw scores ranged from 7.2 on Language Arts Literacy to 8.6 on Mathematics. Raw score to scale score conversion tables by content area are included in Appendix B. Also, frequency distributions of the scale scores by content area are shown in Appendix B.

**TABLE 8.1.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Means and Standard Deviations of Students' Raw Scores by Test Section**

<b>TEST SECTION</b>	<b>Number of Points</b>	<b>Raw Scores Mean</b>	<b>Standard Deviation</b>	<b>Number Tested</b>
<b>Language Arts Literacy</b>	43	23.1	7.2	106,286
<b>Mathematics</b>	42	25.8	8.6	106,134

**Means and Standard Deviations of Students' Raw Scores**

Table 8.1.2 reports the means and standard deviations for students' obtained numbers of raw score points by cluster on the May 2003 test. Table 8.1.2 shows that in Language Arts Literacy, students' mean percent correct was 53.8% overall with 49.6% in Reading and 58.7% in Writing. The mean raw score on the writing/speculate task in response to a picture was 6.0 points out of a possible 10 points and the mean raw score on the writing/analyze task in response to a poem was 5.7 points out of a possible 10 points. The mean percents correct in the two Reading clusters—Working with Text and Analyzing/Critiquing Text—were 64.4% and 41.7%.

With respect to the students' percent correct scores on the Mathematics content clusters, the data in Table 8.1.2 indicate that the mean percent correct ranged from 53.1% in Geometry and Measurement to 66.7% in Number Sense and Numerical Operations. The mathematics items are also categorized as Knowledge (requiring conceptual understanding and procedural knowledge) and Problem Solving. The mean percent correct was 61.4% for Knowledge and 58.3% for Problem Solving.

Table 8.1.3 shows the means and standard deviations for the students' raw scores and percent correct scores on the dichotomously scored items by NJ ASK Content Area. Table 8.1.4 provides means and standard deviations for students' raw scores and percent correct scores on the open-ended items by cluster.

**TABLE 8.1.2**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Means and Standard Deviations of Students' Raw Scores  
and Percent Correct by Content Area**

NJ ASK Content Area	Number of Items		Number of Possible Points	Raw Score		Percent Correct	
	Multiple- Choice	Open- Ended		Mean	Standard Deviation	Mean	Standard Deviation
<b>Language Arts Literacy</b>	11	5	43	23.1	7.2	53.8	16.8
Reading	11	3	23	11.4	4.4	49.6	19.2
Writing	--	2	20	11.7	3.4	58.7	17.2
Writing/Picture	--	1	10	6.0	1.8	60.3	18.3
Writing/Poem	--	1	10	5.7	2.0	57.1	19.8
Working with Text	8	--	8	5.2	2.0	64.4	24.7
Analyzing Text	3	3	15	6.3	2.9	41.7	19.3
<b>Mathematics*</b>	31	5	42	25.8	8.6	61.4	20.5
Number Sense and Numerical Operations*	11	2	13	8.7	3.0	66.7	23.2
Geometry and Measurement	7	1	10	5.3	2.1	53.1	20.7
Patterns and Algebra	6	1	9	5.8	2.2	64.1	24.9
Data Analysis, Probability, and Discrete Math	7	1	10	6.0	2.6	60.3	26.1
Knowledge*	31	5	42	25.8	8.6	61.4	20.5
Problem Solving	17	5	32	18.7	6.9	58.3	21.6

\* Eight multiple-choice items in the Number Sense and Numerical Operations cluster and in the Knowledge skill are counted as one-half point.

**TABLE 8.1.3**

**2003 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**

NJ ASK Content Area	Number of Points	Raw Scores		Percent Correct	
		Mean	Standard Deviation	Mean	Standard Deviation
<b>Language Arts Literacy</b>	11	6.8	2.6	62.3	24.0
Reading	11	6.8	2.6	62.3	24.0
Writing	--	--	--	--	--
Writing/Picture	--	--	--	--	--
Writing/Poem	--	--	--	--	--
Working with Text	8	5.2	2.0	64.4	24.7
Analyzing Text	3	1.7	1.0	56.5	33.7
<b>Mathematics*</b>	27	18.8	5.2	69.6	19.4
Number Sense and Numerical Operations*	7	5.2	1.5	74.8	21.8
Geometry and Measurement	7	4.8	1.7	68.2	23.7
Patterns and Algebra	6	4.4	1.3	73.4	22.5
Data Analysis, Probability and Discrete Math	7	4.4	1.8	62.7	26.4
Knowledge*	27	18.8	5.2	69.6	19.4
Problem Solving	17	11.7	3.5	68.8	20.6

\* Eight items in the Number Sense and Numerical Operations cluster and in the Knowledge skill are counted as one-half point.

**TABLE 8.1.4**

**2003 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 Cluster  
by Content Area**

NJ ASK Content Area	Number		Raw Scores		Percent Correct	
	Items	Points	Mean	Standard Deviation	Mean	Standard Deviation
<b>Language Arts Literacy</b>	5	32	16.3	5.2	51.0	16.3
Reading	3	12	4.6	2.3	38.1	18.8
Writing	2	20	11.7	3.4	58.7	17.2
Writing/Picture	1	10	6.0	1.8	60.3	18.3
Writing/Poem	1	10	5.7	2.0	57.1	19.8
Working with Text	0	0	--	--	--	--
Analyzing	3	12	4.6	2.3	38.0	18.8
<b>Mathematics</b>	5	15	7.0	3.9	46.5	26.0
Number Sense, and Numerical Operations	2	6	3.4	1.9	57.2	31.0
Geometry and Measurement	1	3	0.5	0.8	17.9	26.5
Patterns and Algebra	1	3	1.4	1.3	45.5	42.7
Data Analysis Probability and Discrete Math	1	3	1.6	1.2	54.6	38.8
Knowledge	5	15	7.0	3.9	46.5	26.0
Problem Solving	5	15	7.0	3.9	46.5	26.0

## 8.2 Classical Reliability Estimates of the Test Scores

Table 8.2.1 summarizes reliability estimates for the NJ ASK content areas and clusters. The reliability coefficients given in this table 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 Table 8.2.2.

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 and 8.2.2 reflect the expected positive relationship between test length and reliability.

The SEMs given in Tables 8.2.1 and 8.2.2 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  $\pm 1$  SE of his true score approximately 68 percent of the time, or  $\pm 2$  SE of his true score about 95 percent of the time (p. 252).

**TABLE 8.2.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Reliability Estimates and Standard Errors of Measurement (SEM)  
for Content Areas and Clusters**

<b>NJ ASK Test Section</b>	<b>Number of Points</b>	<b>Reliability</b>	<b>Raw Score SEM</b>	<b>Scale Score SEM</b>
<b>Language Arts Literacy</b>	43	0.85	2.76	9.19
Reading	23	0.82	1.87	
Writing	20	0.77	1.65	
Working with Text	8	0.62	1.21	
Analyzing Text	15	0.76	1.42	
<b>Mathematics</b>	42	0.89	2.81	11.94
Number Sense and Numerical Operations	13	0.78	1.41	
Geometry and Measurement	10	0.60	1.31	
Patterns and Algebra	9	0.54	1.52	
Data analysis, Probability and Discrete Math	10	0.65	1.54	
Knowledge	42	0.89	2.81	
Problem Solving	32	0.86	2.56	

**TABLE 8.2.2****2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Reliability Estimates and Standard Errors of Measurement (SEM)  
for Dichotomously Scored Items Within Content Clusters**

<b>NJ ASK Content Area</b>	<b>Number of Points</b>	<b>Reliability</b>	<b>Raw Score SEM</b>
<b>Language Arts Literacy</b>	11	0.71	1.43
Reading	11	0.71	1.43
Writing*	--	--	--
Writing/Picture	--	--	--
Writing/Poem	--	--	--
<hr/>			
Working with Text	8	0.62	1.21
Analyzing Text	3	0.44	0.75
<b>Mathematics</b>	27	0.85	2.04
Number Sense and Numerical Operations	7	0.72	0.80
Geometry and Measurement	7	0.56	1.09
Patterns and Algebra	6	0.50	0.95
Data analysis, Probability and Discrete Math	7	0.62	1.13
<hr/>			
Knowledge	27	0.85	2.04
Problem Solving	17	0.77	1.67

\* There were no dichotomously scored writing items.

### 8.3 Reliability of Performance Classification

Decision accuracy provides an estimate of how reliably a test form classifies students into performance categories. It 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 Table 8.3.1, the cells showing exact agreement are shaded. The sum of the shaded, diagonal cells represents the estimated proportion correctly classified.

The results of these analyses are presented in Table 8.3.1. For Language Arts Literacy, the estimated proportion correctly classified overall was 0.89. When the decisions were collapsed to below proficient versus proficient and above, the estimated proportion correctly classified was 0.92. 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 0.92.

**TABLE 8.3.1**

**2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Reliability of Classification and Decision Accuracy**

**Decision Accuracy Language Arts Literacy Grade 4**

		<i>Observed Score</i>				
		<b>Placement Score</b>	<b>Advanced Proficient (35-43)</b>	<b>Proficient (18-34.5)</b>	<b>Partially Proficient (0-17.5)</b>	<b>Observed Total</b>
<i>True Score</i>	<b>Advanced Proficient (35-43)</b>	0.00	0.03	0.00	0.04	
	<b>Proficient (18-34.5)</b>	0.00	0.70	0.04	0.74	
	<b>Partially Proficient (0-17.5)</b>	0.00	0.04	0.18	0.22	
	<b>Expected Total</b>	0.00	0.77	0.22		

**Estimated Proportion Correctly Classified: Total = 0.88, Proficient & Above = 0.92**

**Decision Accuracy Mathematics Grade 4**

		<i>Observed Score</i>				
		<b>Placement Score</b>	<b>Advanced Proficient (33-42)</b>	<b>Proficient (22-32.5)</b>	<b>Partially Proficient (0-21.5)</b>	<b>Observed Total</b>
<i>True Score</i>	<b>Advanced Proficient (33-42)</b>	0.20	0.05	0.00	0.25	
	<b>Proficient (22-32.5)</b>	0.04	0.35	0.04	0.43	
	<b>Partially Proficient (0-21.5)</b>	0.00	0.04	0.28	0.32	
	<b>Expected Total</b>	0.24	0.44	0.32		

**Estimated Proportion Correctly Classified: Total = 0.83, Proficient & Above = 0.92**

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

**TABLE 8.4.1**

**2003 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
4	LAL	Proficient	18.0	-0.0303	0.1916	2.5
		Advanced Proficient	35.0	1.4277	0.2377	2.0
	Math	Proficient	22.0	0.4848	0.1665	3.0
		Advanced Proficient	33.0	1.1582	0.2016	2.5

## 8.5 Rater Reliability

Table 8.5.1 shows 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 May 2003, 57.2% received exactly the same scores by the raters and 39.6% received scores that were adjacent. Thus, a total of 96.8% of the task responses required only two raters. The remaining 3.2% received scores on the Writing Tasks that differed by more than one point and therefore required a third rater.

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.

For the three reading open-ended items, the resolution percent ranges from 2.9% to 4.3% with the percent at perfect agreement ranging from 55.4% to 58.3%.

One open-ended item was presented for each of the five Mathematics clusters. These five mathematics items had percents at perfect agreement ranging from 75.8% to 88.0%. The percent requiring resolution ranged from 0.7% to 2.8%.

**TABLE 8.5.1****2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Consistency Between Raters Scoring Writing Tasks and Open-Ended Items**

<b>Writing Tasks and Open-Ended Items</b>	<b>Percent Raters In Exact Agreement</b>	<b>Percent Raters In Adjacent Agreement</b>	<b>Percent Resolution Needed</b>
<b>Language Arts Literacy</b>	56.8	39.6	3.5
<b>Writing Total</b>	<b>57.2</b>	<b>39.6</b>	<b>3.2</b>
Writing/Picture	53.9	41.7	4.3
Writing/Poem	60.5	37.4	2.1
<b>Reading Total</b>	<b>56.6</b>	<b>39.7</b>	<b>3.8</b>
Open-Ended Item 1	58.3	38.9	2.9
Open-Ended Item 2	55.4	40.5	4.1
Open-Ended Item 3	56.0	39.7	4.3
<b>Mathematics</b>	<b>82.8</b>	<b>15.6</b>	<b>1.6</b>
Open-Ended Item 1	75.8	23.1	1.2
Open-Ended Item 2	79.0	18.1	2.8
Open-Ended Item 3	88.0	11.1	0.9
Open-Ended Item 4	83.6	14.2	2.2
Open-Ended Item 5	87.5	11.8	0.7

## **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. As the NJASK3 was administered as a field test in 2003, Standard Setting was not held for this test and scale scores and proficiency level information is not available. Therefore,

for the 2003 administration, the reports for the NJASK3 will differ slightly from those for the NJASK4. NJASK3 reports will not show scaled scores, proficiency levels, or individual cluster level totals. The NJASK3 reports will show total raw scores. 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

2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
Individual Student Report (ISR) – Front

# Individual Student Report

USING ASSESSMENT TO HELP STUDENTS GROW

NEW JERSEY  
ASSESSMENT OF SKILLS  
AND KNOWLEDGE  
SPRING 2003

*This report contains information from the Spring 2003 administration of the New Jersey Assessment of Skills and Knowledge.*

*In this report you will find:*

- Your child's overall score and overall proficiency level in Language Arts Literacy and Mathematics
- Your child's performance in each of the content clusters on the tests

*Please remember that a single test can provide only limited information. To learn more, review your child's report card and talk to your child's teacher.*

**John M. Allen**

Gender: Male  
Date of Birth: 02/12/94  
Grade level of test: 4

School: 339 Madison Elementary  
District: 3393 Adams Regional  
School District  
County: 33 Monroe County

### Language Arts Literacy

Your child's scale score is 195, which is at the Partially Proficient level.

Your child has not met the New Jersey standard of 200 or above in Language Arts Literacy. Reading with your child daily will help your child develop and improve in language arts.

### Mathematics

Your child's scale score is 248, which is at the Proficient level.

Your child has met the New Jersey standard of 200 or above in Mathematics. To help your child continue to develop and improve in mathematics, encourage your child to apply math concepts to daily activities.

Year class score: 195

Year class score: 248

### About Scale Scores and Proficiency Levels

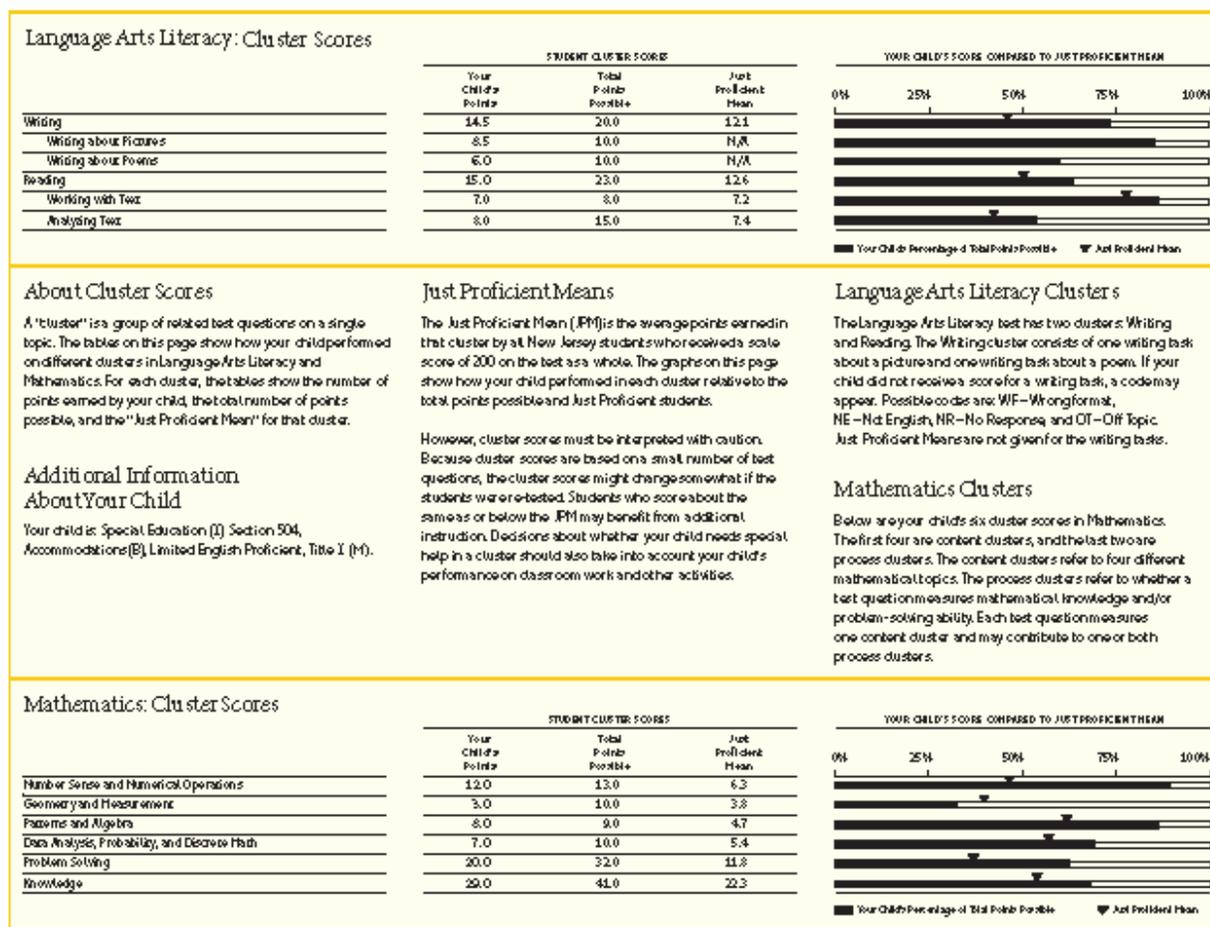
Your child's Language Arts Literacy and Mathematics scale scores are presented in the box on the left. The scale scores are based on the number of correct answers to multiple-choice questions and the number of points earned for writing samples and open-ended questions.

The possible scale scores for each subject are from 100 to 300. If the scale score is below 200, your child scored "Partially Proficient" in that subject. If the scale score is at or above 200 but below 250, your child scored "Proficient" in that subject. If the scale score is at or above 250, your child scored "Advanced Proficient" in that subject.

Under the standards set by New Jersey, students are expected to perform at the Proficient level or above.

Figure 10.1.2

2003 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 IEP Exempt, who did not take 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 IEP Exempt, who did not take 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, limited English proficient, and Title I 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, limited English proficient, and Title I students combined. The report format is the same as the summary of school performance. Any district that chooses to test a student whose IEP exempts them 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, and limited English proficient 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, and limited English proficient students combined. The report format is the same as the summary of school cluster performance. Any district that chooses to test a student whose IEP exempts them 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, and Statewide Disaggregated Student Population 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 is divided into two sections. The top section provides more detailed test score information for total (all students), general education, special education, limited English proficient, and Title I students. The bottom section of the page provides cluster raw score information.

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 – Group Performance) and Figure 10.2.2 (back page – Cluster Performance).

Figure 10.2.1

2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
School Report - Performance by Demographic Groups



County: 00 Monroe County  
District: 0000 Adams Regional School District  
School: 000 Madison Elementary

Group Performance

	Number of Students Enrolled <sup>1</sup>	Number of Students Tested	Number of Valid Scale Scores	Scale Score Mean	% Partially Proficient	% Proficient	% Advanced Proficient
ALL Students	373	372	371	257.2	10.5%	42.3%	47.2%
General Education <sup>2</sup>	273	273	273	259.1	12.2%	59.2%	28.6%
Special Education	100	99	98	249.1	23.6%	54.4%	22.0%
Limited English Proficient	0	0	0	—	—	—	—
<b>Total</b>	<b>50</b>	<b>50</b>	<b>49</b>	<b>251.2</b>	<b>4.0%</b>	<b>62.0%</b>	<b>34.0%</b>
<b>Gender<sup>2</sup></b>							
Female	170	169	168	244.6	6.0%	39.9%	55.1%
Male	203	202	201	247.5	5.5%	35.5%	59.0%
<b>Ethnicity<sup>2</sup></b>							
American Indian	0	0	0	—	—	—	—
Asian	50	50	49	253.6	6.0%	24.0%	70.0%
Black	65	63	62	246.3	5.7%	36.9%	57.4%
Hispanic	18	18	17	253.6	6.0%	24.0%	70.0%
Pacific Islander	0	0	0	—	—	—	—
White	209	209	208	245.0	5.6%	39.7%	54.6%
Other	23	23	22	231.1	4.0%	75.8%	20.2%
Multiple	10	10	9	253.6	6.0%	24.0%	70.0%
<b>Economic Status<sup>2</sup></b>							
Economically Disadvantaged	1	1	1	182.0	100.0%	0.0%	0.0%
Non-Economically Disadvantaged	372	372	370	246.4	5.7%	36.8%	57.6%
<b>Migrant Status<sup>2</sup></b>							
Migrant	0	0	0	—	—	—	—
Non-Migrant	373	371	370	257.2	5.7%	36.9%	57.4%

Note: Percentages might not total 100 due to rounding.

— No students in this category.

1. Enrollment is based on the number of available test booklets.

2. General Education excludes Special Education and Limited English Proficient students.

3. Possible differences in total among demographic categories may have resulted from gridding errors or missing data in raw data received from districts.

See page 2  
for user information

03/03/03

Page 1 of 2

Figure 10.2.2

2003 New Jersey Assessment of Skills and Knowledge (NJ ASK)  
School Report - Cluster Score Means

SCHOOL REPORT: MATHEMATICS  
SPRING 2003 • CYCLE 31 • GRADE 4

County: 00 Monroe County
District: 0000 Adams Regional School District
School: 000 Hudson Elementary

Cluster Performance

	Total Points Possible	Just Proficient Mean (JPF) <sup>1</sup>	School Mean	District Mean	DEGI Mean	State Mean
<b>1. Number Sense &amp; Numerical Operations</b>						
All Students	13	6.3	6.9	7.1	6.9	6.5
General Education <sup>2</sup>			6.9	7.1	6.9	6.5
Special Education			5.3	5.2	4.9	4.6
Limited English Proficient			—	5.7	5.2	5.0
<b>2. Geometry &amp; Measurement</b>						
All Students	10	3.8	3.3	3.8	3.4	3.8
General Education <sup>2</sup>			4.6	4.7	4.3	3.8
Special Education			3.9	3.7	3.9	3.3
Limited English Proficient			—	3.7	3.5	3.1
<b>3. Patterns &amp; Algebra</b>						
All Students	9	4.7	6.3	6.2	5.8	5.8
General Education <sup>2</sup>			6.4	6.1	5.9	5.4
Special Education			6.3	6.6	5.9	5.2
Limited English Proficient			—	6.1	6.4	5.1
<b>4. Data Analysis, Probability, &amp; Discrete Math</b>						
All Students	10	5.4	3.3	3.8	3.4	3.8
General Education <sup>2</sup>			6.8	6.8	6.5	5.7
Special Education			6.1	6.2	5.9	5.6
Limited English Proficient			—	4.0	6.4	5.6
<b>5. Problem Solving</b>						
All Students	32	11.8	14.3	13.8	12.4	12.1
General Education <sup>2</sup>			15.8	15.1	13.9	12.8
Special Education			14.3	13.8	12.4	12.1
Limited English Proficient			—	11.6	11.8	11.2
<b>6. Knowledge</b>						
All Students	41	22.3	20.1	20.0	19.8	19.6
General Education <sup>2</sup>			20.2	20.1	19.9	19.7
Special Education			20.1	19.9	19.9	19.4
Limited English Proficient			—	19.3	19.2	19.1

— No students in this category.

<sup>1</sup> The numbers in this column are the scaled mean score based on students whose scaled score is 200.

<sup>2</sup> General Education excludes Special Education and Limited English Proficient students.

03/03/03

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.

### **10.3 Interpreting Reports**

The 2003 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.

### **10.4 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:**

### **State Summary Executive Summary**

The spring 2003 New Jersey Assessment of Skills and Knowledge (NJ ASK) consists of two sections: 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 for the two content areas. 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 either 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 May 19 and May 30, 2003. The Language Arts Literacy test was administered to 106,286 total students and Mathematics was administered to 106,134 total students.

This executive summary includes a series of tables summarizing test results for the State, Special Needs Districts, All Other (Non Special Needs) Districts, and District Factor Groups. The tables that follow are derived from the statewide performance data of the Cycle II report. Table A.1 presents statewide results for Language Arts Literacy and Table A.2 presents statewide results for Mathematics. Results for Tables A.1 and A.2 are presented for the following student groups: all, 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, tested, and with valid scale scores. Enrollment is based on the number of scannable test booklets. 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 percentages may not total to one hundred due to rounding. An explanation of the District Factor Groups (DFGs) is included at the end of this report. It should be noted that results reported at the State level include some students whose scores are not reflected at the DFG or charter school level.

The NJ ASK scores in this report exclude students whose tests were voided for a particular test section. The data in this report are based on information collected from data submitted on students' tests. Subsequent to the initial distribution of test results in August 2003, a process was completed to correct errors in student information made when students' tests were originally

submitted by the district. Those corrections to student information received from districts by the State Department of Education by December 2003 have been included in these summaries.

Following are highlights of the 2003 assessment results.

- Of the 106,286 students who attempted the Language Arts Literacy section in Spring 2003, 22.4% scored in Partially Proficient; 73.8% scored in Proficient; and 3.8% scored in Advanced Proficient. (Table A.1)
- The mean scale score on the Language Arts Literacy section was 214.6 in Spring 2003. (Table A.1)
- Of the 106,134 students who attempted the Mathematics section in Spring 2003, 32.0% scored Partially Proficient; 42.8% scored in Proficient; and 25.2% scored in Advanced Proficient. (Table A.2)
- The mean scale score on the Mathematics section was 217.3 in Spring 2003. (Table A.2)
- Of the students who attempted the Language Arts Literacy section, the percents in Partially Proficient ranged from 45.3% in DFG A to 5.8% in DFG J. In Proficient, the percents ranged from 54.1% in DFG A to 85.1% in DFG J. The percents in Advanced Proficient ranged from 0.6% in DFG A to 9.1% in DFG J. (Table A.3 – Total Students)
- Of the students who attempted the Mathematics section, the percents in Partially Proficient ranged from 55.4% in DFG A to 10.6% in DFG J. In Proficient, the percents ranged from 33.7% in DFG A to 47.1% in DFG FG. The percents in Advanced Proficient ranged from 11.0% in DFG A to 44.2% in DFG J. (Table A.4 – Total Students)
- In Language Arts Literacy, 44.2% of the students in the Special Needs Districts scored in Partially Proficient; 55.1% scored in Proficient; and 0.7% scored in Advanced Proficient. (Table A.5)
- In Mathematics, 55.0% of the students in the Special Needs Districts scored in Partially Proficient; 34.0% scored in Proficient; and 11.1% scored in Advanced Proficient. (Table A.5)

**TABLE A.1**

**NJ ASK Statewide Performance by Demographic Groups Spring 2003  
Language Arts Literacy**

		<b>Number of Students Enrolled</b>	<b>Number of Students Tested</b>	<b>Number of Valid Scores</b>	<b>Scale Score Mean</b>	<b>% Partially Proficient</b>	<b>% Proficient</b>	<b>% Advanced Proficient</b>
<b>All</b>	All Students	107345	106957	106286	214.6	22.4%	73.8%	3.8%
<b>Education Status</b>	General Education	87122	87086	86744	220.1	13.9%	81.5%	4.6%
	Special Education	16252	15900	15687	191.9	58.5%	41.1%	0.4%
<b>Limited English Proficient</b>	Limited English Proficient	4237	4234	4111	183.0	68.6%	31.4%	0.0%
<b>Title</b>	Title I	20570	20514	20326	200.2	43.1%	56.4%	0.6%
<b>Gender</b>	Female	52201	52068	51820	219.1	16.8%	77.6%	5.6%
	Male	54985	54740	54330	210.3	27.7%	70.2%	2.1%
<b>Ethnicity</b>	American Indian	112	112	112	213.8	21.4%	76.8%	1.8%
	Asian	6497	6481	6450	224.9	10.6%	80.9%	8.4%
	Black	19510	19415	19224	201.8	42.0%	57.1%	0.8%
	Hispanic	17712	17638	17377	203.1	37.5%	61.4%	1.1%
	Pacific Islander	373	373	372	222.3	10.8%	81.7%	7.5%
	White	60664	60483	60327	220.9	13.1%	81.8%	5.1%
	Other	441	440	437	214.1	22.4%	73.0%	4.6%
	Multiple	838	835	831	213.1	24.1%	72.6%	3.4%
<b>Economic Status</b>	Economically Disadvantaged	31962	31835	31458	201.1	41.8%	57.6%	0.6%
	Non-Economically Disadvantaged	75383	75122	74828	220.3	14.2%	80.6%	5.2%
<b>Migrant Status</b>	Migrant	75	74	74	189.0	63.5%	35.1%	1.4%
	Non-Migrant	107270	106883	106212	214.6	22.4%	73.8%	3.8%

**TABLE A.2**

**NJ ASK Statewide Performance by Demographic Groups Spring 2003  
Mathematics**

		<b>Number of Students Enrolled</b>	<b>Number of Students Tested</b>	<b>Number of Valid Scores</b>	<b>Scale Score Mean</b>	<b>% Partially Proficient</b>	<b>% Proficient</b>	<b>% Advanced Proficient</b>
<b>All</b>	All Students	106969	106803	106134	217.3	32.0%	42.8%	25.2%
<b>Education Status</b>	General Education	86998	86976	86596	223.4	25.2%	45.8%	29.0%
	Special Education	16001	15857	15658	190.4	61.6%	30.0%	8.4%
<b>Limited English Proficient</b>	Limited English Proficient	4231	4231	4137	188.0	66.0%	26.8%	7.2%
<b>Title</b>	Title I	20039	20024	19837	194.3	59.2%	32.1%	8.7%
<b>Gender</b>	Female	52061	52002	51747	215.8	33.4%	43.4%	23.2%
	Male	54759	54658	54253	218.8	30.7%	42.2%	27.1%
<b>Ethnicity</b>	American Indian	112	112	112	216.3	32.1%	42.0%	25.9%
	Asian	6478	6471	6441	238.3	12.9%	39.5%	47.6%
	Black	19436	19400	19203	194.3	58.1%	33.4%	8.5%
	Hispanic	17631	17608	17377	202.9	47.6%	39.2%	13.2%
	Pacific Islander	373	373	373	230.8	18.0%	46.6%	35.4%
	White	60485	60393	60205	226.5	21.4%	47.1%	31.5%
	Other	440	440	438	216.1	32.0%	47.3%	20.8%
	Multiple	835	833	828	216.0	33.7%	40.9%	25.4%
<b>Economic Status</b>	Economically Disadvantaged	31844	31799	31449	198.3	53.4%	35.8%	10.8%
	Non-Economically Disadvantaged	75125	75004	74685	225.3	23.0%	45.7%	31.3%
<b>Migrant Status</b>	Migrant	75	74	74	190.9	60.8%	29.7%	9.5%
	Non-Migrant	106894	106729	106060	217.3	32.0%	42.8%	25.2%



**TABLE A.3**

**New Jersey Statewide Testing System  
Spring 2003 New Jersey Assessment of Skills and Knowledge  
By District Factor Group  
Language Arts Literacy**

**GENERAL EDUCATION STUDENTS <sup>b</sup>**

DFG	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	14,807	65.8	34.2	65.0	0.8	206.3
B	9,129	80.0	20.0	78.3	1.7	214.1
CD	7,442	84.5	15.5	82.6	1.9	216.9
DE	12,817	90.0	10.0	86.4	3.6	221.3
FG	10,946	91.7	8.3	86.8	5.0	223.3
GH	12,105	93.5	6.5	87.3	6.2	225.4
I	16,757	96.8	3.2	87.1	9.8	230.3
J	1,588	97.4	2.6	87.1	10.3	231.2

**SPECIAL EDUCATION STUDENTS <sup>c</sup>**

DFG	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	2,929	18.3	81.7	18.3	0.0	175.8
B	1,791	28.0	72.0	28.0	0.1	185.0
CD	1,580	32.1	67.9	32.1	0.0	188.1
DE	2,328	43.5	56.5	43.2	0.3	193.4
FG	1,863	46.1	53.9	45.7	0.4	195.6
GH	2,055	54.2	45.8	53.4	0.8	200.1
I	2,710	65.5	34.5	64.4	1.2	205.6
J	214	73.8	26.2	73.4	0.5	208.6

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND IEP 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 A.3 (continued)**

**New Jersey Statewide Testing System  
Spring 2003 New Jersey Assessment of Skills and Knowledge  
By District Factor Group**

**Language Arts Literacy Section**

**LIMITED ENGLISH PROFICIENT STUDENTS <sup>d</sup>**

DFG	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	2,353	26.9	73.1	26.9	0.0	179.5
B	515	24.3	75.7	24.3	0.0	178.6
CD	250	28.4	71.6	28.4	0.0	183.8
DE	302	44.7	55.3	44.7	0.0	191.5
FG	206	39.8	60.2	39.8	0.0	191.7
GH	221	43.0	57.0	43.0	0.0	190.8
I	237	59.9	40.1	59.5	0.4	200.1
J	13	38.5	61.5	38.5	0.0	190.5

**TOTAL STUDENTS <sup>e</sup>**

DFG	VALID SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	19,915	54.7	45.3	54.1	0.6	198.9
B	11,414	69.4	30.6	68.1	1.4	208.0
CD	9,256	74.2	25.8	72.6	1.6	211.1
DE	15,431	82.2	17.8	79.1	3.1	216.5
FG	13,012	84.4	15.6	80.2	4.2	218.8
GH	14,372	87.2	12.8	81.8	5.4	221.3
I	19,695	92.1	7.9	83.6	8.5	226.6
J	1,814	94.2	5.8	85.1	9.1	228.3

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND IEP 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 A.3 (continued)**

**New Jersey Statewide Testing System  
Spring 2003 New Jersey Assessment of Skills and Knowledge  
Language Arts Literacy Section**

**CHARTER SCHOOLS <sup>f</sup>**

	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL <sup>b</sup> EDUCATION STUDENTS	1,143	61.5	38.5	60.3	1.2	204.7
SPECIAL <sup>c</sup> EDUCATION STUDENTS	105	28.6	71.4	28.6	0.0	186.5
LIMITED ENGLISH <sup>d</sup> PROFICIENT STUDENTS	7	28.6	71.4	28.6	0.0	183.7
TOTAL <sup>e</sup> STUDENTS	1,255	58.6	41.4	57.5	1.1	203.1

**STATEWIDE RESULTS**

	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL <sup>b</sup> EDUCATION STUDENTS	86,744	86.1	13.9	81.5	4.6	220.1
SPECIAL <sup>c</sup> EDUCATION STUDENTS	15,687	41.5	58.5	41.1	0.4	191.9
LIMITED ENGLISH <sup>d</sup> PROFICIENT STUDENTS	4,111	31.4	68.6	31.4	0.0	183.0
TOTAL <sup>e</sup> STUDENTS	106,286	77.6	22.4	73.8	3.8	214.6

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND IEP 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 A.4**

**New Jersey Statewide Testing System  
Spring 2003 New Jersey Assessment of Skills and Knowledge  
By District Factor Group  
Mathematics Section**

**GENERAL EDUCATION STUDENTS <sup>b</sup>**

DFG	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	14,807	51.7	48.3	38.3	13.4	203.2
B	9,121	65.6	34.4	46.2	19.4	214.4
CD	7,439	72.0	28.0	49.4	22.7	219.4
DE	12,775	78.0	22.0	49.0	29.0	225.3
FG	10,938	80.5	19.5	49.8	30.8	227.6
GH	12,029	83.3	16.7	47.0	36.3	231.4
I	16,746	89.8	10.2	45.8	44.0	238.4
J	1,587	92.7	7.3	44.5	48.1	241.7

**SPECIAL EDUCATION STUDENTS <sup>c</sup>**

DFG	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	2,922	20.1	79.9	16.8	3.2	173.0
B	1,794	28.4	71.6	23.6	4.7	182.2
CD	1,579	32.7	67.3	27.0	5.8	186.4
DE	2,318	37.7	62.3	29.9	7.8	191.0
FG	1,862	41.7	58.3	32.8	9.0	193.3
GH	2,043	49.3	50.7	37.9	11.4	200.0
I	2,710	58.0	42.0	42.1	15.9	207.2
J	215	66.0	34.0	49.3	16.7	211.3

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND IEP 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 A.4 (continued)**

**New Jersey Statewide Testing System  
Spring 2003 New Jersey Assessment of Skills and Knowledge  
By District Factor Group  
Mathematics Section**

**LIMITED ENGLISH PROFICIENT STUDENTS <sup>d</sup>**

DFG	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	2,369	28.8	71.2	24.3	4.5	183.5
B	518	27.0	73.0	21.6	5.4	182.2
CD	252	40.1	59.9	30.6	9.5	192.6
DE	304	43.4	56.6	33.6	9.9	196.3
FG	207	43.5	56.5	32.4	11.1	196.0
GH	221	42.1	57.9	31.7	10.4	195.7
I	239	65.3	34.7	39.7	25.5	215.2
J	13	61.5	38.5	53.8	7.7	201.3

**TOTAL STUDENTS <sup>e</sup>**

DFG	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
A	19,923	44.6	55.4	33.7	11.0	196.7
B	11,412	58.1	41.9	41.6	16.5	208.0
CD	9,254	64.5	35.5	45.1	19.5	213.1
DE	15,381	71.3	28.7	45.9	25.4	219.6
FG	13,004	74.4	25.6	47.1	27.3	222.2
GH	14,284	77.8	22.2	45.5	32.4	226.4
I	19,686	85.1	14.9	45.2	39.9	233.8
J	1,814	89.4	10.6	45.2	44.2	237.9

a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND IEP 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 A.4 (continued)**

**New Jersey Statewide Testing System  
Spring 2003 New Jersey Assessment of Skills and Knowledge  
Mathematics Section**

**CHARTER SCHOOLS <sup>f</sup>**

	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL <sup>b</sup> EDUCATION STUDENTS	1,144	41.5	58.5	31.1	10.4	195.1
SPECIAL <sup>c</sup> EDUCATION STUDENTS	105	21.0	79.0	18.1	2.9	178.2
LIMITED ENGLISH <sup>d</sup> PROFICIENT STUDENTS	7	28.6	71.4	28.6	0.0	196.6
TOTAL <sup>e</sup> STUDENTS	1,256	39.7	60.3	30.0	9.7	193.7

**STATEWIDE RESULTS**

	VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
			PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL <sup>b</sup> EDUCATION STUDENTS	86,596	74.8	25.2	45.8	29.0	223.4
SPECIAL <sup>c</sup> EDUCATION STUDENTS	15,658	38.4	61.6	30.0	8.4	190.4
LIMITED ENGLISH <sup>d</sup> PROFICIENT STUDENTS	4,137	34.0	66.0	26.8	7.2	188.0
TOTAL <sup>e</sup> STUDENTS	106,134	68.0	32.0	42.8	25.2	217.3

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND IEP 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 A.5**  
**New Jersey Statewide Testing System**  
**Spring 2003 New Jersey Assessment of Skills and Knowledge**  
**Language Arts Literacy and Mathematics for**  
**The Special Needs Districts as Compared to All Other Districts**

LANGUAGE ARTS LITERACY SECTION		VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
				PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL <sup>b</sup> EDUCATION STUDENTS	SPECIAL NEEDS	16,931	67.1	32.9	66.2	0.9	206.9
	ALL OTHERS	69,813	90.7	9.3	85.2	5.5	223.3
SPECIAL <sup>c</sup> EDUCATION STUDENTS	SPECIAL NEEDS	3,386	18.0	82.0	18.0	0.0	176.4
	ALL OTHERS	12,301	48.0	52.0	47.5	0.5	196.1
LIMITED ENGLISH <sup>d</sup> PROFICIENT STUDENTS	SPECIAL NEEDS	2,459	26.5	73.5	26.5	0.0	179.4
	ALL OTHERS	1,652	38.6	61.4	38.6	0.1	188.3
TOTAL <sup>e</sup> STUDENTS	SPECIAL NEEDS	22,602	55.8	44.2	55.1	0.7	199.6
	ALL OTHERS	83,684	83.5	16.5	78.8	4.7	218.7

MATHEMATICS SECTION		VALID <sup>a</sup> SCORES 2003	PERCENT WHO SCORED ADVANCED PROFICIENT OR PROFICIENT 2003	PERCENT AT EACH PROFICIENCY LEVELS			MEAN SCALE SCORE 2003
				PARTIALLY PROFICIENT (100-199)	PROFICIENT (200-249)	ADVANCED PROFICIENT (250-300)	
GENERAL <sup>b</sup> EDUCATION STUDENTS	SPECIAL NEEDS	16,924	52.3	47.7	38.8	13.5	203.7
	ALL OTHERS	69,672	80.3	19.7	47.5	32.8	228.2
SPECIAL <sup>c</sup> EDUCATION STUDENTS	SPECIAL NEEDS	3,377	19.7	80.3	16.4	3.3	173.0
	ALL OTHERS	12,281	43.6	56.4	33.7	9.9	195.2
LIMITED ENGLISH <sup>d</sup> PROFICIENT STUDENTS	SPECIAL NEEDS	2,476	27.9	72.1	23.6	4.3	182.6
	ALL OTHERS	1,661	43.0	57.0	31.5	11.5	196.1
TOTAL <sup>e</sup> STUDENTS	SPECIAL NEEDS	22,602	45.0	55.0	34.0	11.1	197.1
	ALL OTHERS	83,532	74.2	25.8	45.2	29.0	222.8

- a. EXCLUDES STUDENTS' TEST BOOKLETS CODED VOID AND IEP 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.

The **“No. Tested”** column excludes students’ test booklets coded Void. If the number of students tested for a particular group was less than or equal to 10, no summary data are reported. Additionally, if the number of students tested in general education, special education or limited English proficient is equal to 1, and data for each of the other groups are displayed, no summary data are reported for “Total.”

## APPENDIX B:

### 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. Percent of adult residents who failed to complete high school
- B. Percent 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: percent of those in the work force who received some unemployment compensation
- G. Poverty: percent 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 Score – Scale Score Conversions with Frequencies**  
**2003 NJ ASK Language Arts Literacy**

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students	Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
0	107	-2.6910	1.460	45	0	22	213	0.2696	0.196	44598	42.0
0.5	110	-2.0053	0.706	48	0	22.5	214	0.3083	0.197	47366	44.6
1	114	-1.6847	0.464	118	0.1	23	216	0.3475	0.198	50140	47.2
1.5	117	-1.5149	0.369	128	0.1	23.5	218	0.3870	0.199	53049	49.9
2	120	-1.3984	0.317	257	0.2	24	219	0.4269	0.200	55965	52.7
2.5	124	-1.3080	0.285	270	0.3	24.5	221	0.4673	0.201	58822	55.3
3	128	-1.2329	0.263	446	0.4	25	222	0.5080	0.202	61750	58.1
3.5	131	-1.1678	0.247	480	0.5	25.5	224	0.5491	0.203	64682	60.9
4	135	-1.1096	0.235	695	0.7	26	225	0.5907	0.204	67770	63.8
4.5	139	-1.0565	0.225	762	0.7	26.5	227	0.6325	0.205	70642	66.5
5	143	-1.0074	0.217	1069	1.0	27	228	0.6749	0.206	73615	69.3
5.5	146	-0.9613	0.211	1195	1.1	27.5	229	0.7175	0.207	76415	71.9
6	149	-0.9176	0.206	1588	1.5	28	231	0.7605	0.207	79247	74.6
6.5	152	-0.8760	0.202	1776	1.7	28.5	232	0.8039	0.208	81809	77.0
7	155	-0.8358	0.198	2304	2.2	29	233	0.8477	0.209	84275	79.3
7.5	158	-0.7969	0.195	2585	2.4	29.5	235	0.8919	0.210	86697	81.6
8	160	-0.7590	0.193	3207	3.0	30	236	0.9365	0.211	88933	83.7
8.5	163	-0.7217	0.192	3610	3.4	30.5	237	0.9816	0.212	90948	85.6
9	165	-0.6849	0.191	4320	4.1	31	239	1.0272	0.214	92902	87.4
9.5	167	-0.6484	0.190	4815	4.5	31.5	240	1.0734	0.215	94763	89.2
10	169	-0.6123	0.190	5573	5.2	32	241	1.1204	0.217	96399	90.7
10.5	171	-0.5761	0.190	6174	5.8	32.5	243	1.1683	0.219	97796	92.0
11	173	-0.5400	0.190	7046	6.6	33	244	1.2173	0.222	99130	93.3
11.5	175	-0.5038	0.190	7807	7.3	33.5	245	1.2674	0.225	100285	94.4
12	177	-0.4676	0.190	8750	8.2	34	247	1.3190	0.228	101287	95.3
12.5	179	-0.4314	0.190	9616	9.0	34.5	248	1.3723	0.233	102220	96.2
13	181	-0.3951	0.190	10757	10.1	35	250	1.4277	0.237	102988	96.9
13.5	183	-0.3587	0.190	11792	11.1	35.5	251	1.4854	0.243	103683	97.6
14	185	-0.3223	0.190	13093	12.3	36	252	1.5462	0.249	104286	98.1
14.5	187	-0.2859	0.190	14332	13.5	36.5	254	1.6103	0.257	104790	98.6
15	189	-0.2495	0.190	15789	14.9	37	256	1.6786	0.265	105152	98.9
15.5	190	-0.2130	0.190	17174	16.2	37.5	258	1.7520	0.276	105450	99.2
16	192	-0.1766	0.191	18737	17.6	38	259	1.8316	0.288	105683	99.4
16.5	194	-0.1401	0.191	20311	19.1	38.5	262	1.9189	0.303	105860	99.6
17	196	-0.1035	0.191	22195	20.9	39	264	2.0159	0.320	106003	99.7
17.5	198	-0.0670	0.191	23802	22.4	39.5	267	2.1257	0.342	106112	99.8
18	200	-0.0303	0.191	25739	24.2	40	269	2.2521	0.369	106190	99.9
18.5	201	0.0065	0.191	27711	26.1	40.5	273	2.4021	0.405	106226	99.9
19	203	0.0434	0.192	29969	28.2	41	277	2.5864	0.455	106260	100
19.5	205	0.0805	0.192	32145	30.2	41.5	281	2.8260	0.528	106275	100
20	206	0.1178	0.193	34516	32.5	42	285	3.1688	0.653	106284	100
20.5	208	0.1553	0.194	36885	34.7	42.5	292	3.7720	0.942	106285	100
21	210	0.1931	0.194	39428	37.1	43	294	4.8782	1.776	106286	100
21.5	211	0.2312	0.195	41938	39.5						

## 2003 NJ ASK Mathematics

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
0	104	-4.4144	1.859	0	0
0.5	106	-3.1419	1.035	2	0
1	108	-2.4061	0.720	9	0
1.5	111	-1.9965	0.571	17	0
2	114	-1.7216	0.482	29	0
2.5	117	-1.5176	0.423	49	0
3	119	-1.3561	0.381	91	0.1
3.5	122	-1.2224	0.350	141	0.1
4	125	-1.1081	0.326	225	0.2
4.5	127	-1.0080	0.307	317	0.3
5	130	-0.9187	0.291	468	0.4
5.5	132	-0.8379	0.277	679	0.6
6	135	-0.7639	0.266	928	0.9
6.5	137	-0.6956	0.256	1235	1.2
7	139	-0.6320	0.248	1588	1.5
7.5	142	-0.5723	0.240	2027	1.9
8	144	-0.5161	0.233	2533	2.4
8.5	146	-0.4629	0.227	3118	2.9
9	148	-0.4123	0.222	3706	3.5
9.5	151	-0.3640	0.217	4402	4.1
10	153	-0.3178	0.212	5153	4.9
10.5	155	-0.2733	0.208	5977	5.6
11	157	-0.2306	0.204	6862	6.5
11.5	159	-0.1894	0.201	7764	7.3
12	161	-0.1495	0.198	8666	8.2
12.5	163	-0.1109	0.195	9666	9.1
13	165	-0.0735	0.192	10712	10.1
13.5	167	-0.0370	0.189	11710	11.0
14	169	-0.0015	0.187	12814	12.1
14.5	171	0.0330	0.184	13962	13.2
15	173	0.0668	0.182	15169	14.3
15.5	175	0.0999	0.180	16355	15.4
16	177	0.1322	0.179	17622	16.6
16.5	179	0.1639	0.177	18934	17.8
17	180	0.1951	0.175	20318	19.1
17.5	182	0.2257	0.174	21680	20.4
18	184	0.2559	0.173	23094	21.8
18.5	186	0.2856	0.171	24569	23.1
19	188	0.3149	0.170	26076	24.6
19.5	190	0.3439	0.169	27566	26.0
20	192	0.3725	0.168	29152	27.5
20.5	194	0.4009	0.168	30696	28.9
21	196	0.4291	0.167	32371	30.5

Raw Score	Scale Score	Theta	S.E.	Cumulative Number of Students	Cumulative Percent of Students
21.5	198	0.4570	0.166	33985	32.0
22	200	0.4848	0.166	35709	33.6
22.5	201	0.5125	0.166	37458	35.3
23	204	0.5401	0.166	39293	37.0
23.5	206	0.5676	0.165	41095	38.7
24	208	0.5951	0.165	43017	40.5
24.5	210	0.6226	0.166	44980	42.4
25	212	0.6503	0.166	46909	44.2
25.5	214	0.6780	0.166	48870	46.0
26	216	0.7059	0.167	50895	48.0
26.5	218	0.7341	0.168	52975	49.9
27	220	0.7624	0.169	55031	51.9
27.5	222	0.7912	0.170	57112	53.8
28	225	0.8204	0.171	59344	55.9
28.5	227	0.8500	0.173	61457	57.9
29	229	0.8803	0.174	63713	60.0
29.5	232	0.9113	0.177	65898	62.1
30	234	0.9430	0.179	68240	64.3
30.5	237	0.9757	0.182	70379	66.3
31	239	1.0094	0.185	72692	68.5
31.5	241	1.0443	0.188	74846	70.5
32	244	1.0807	0.192	77226	72.8
32.5	246	1.1186	0.196	79399	74.8
33	250	1.1582	0.201	81728	77.0
33.5	251	1.1999	0.206	83913	79.1
34	254	1.2438	0.212	86236	81.3
34.5	256	1.2902	0.218	88272	83.2
35	259	1.3393	0.224	90616	85.4
35.5	261	1.3913	0.231	92562	87.2
36	263	1.4466	0.238	94621	89.2
36.5	266	1.5054	0.246	96413	90.8
37	268	1.5680	0.254	98300	92.6
37.5	269	1.6349	0.263	99814	94.0
38	271	1.7068	0.273	101405	95.5
38.5	273	1.7850	0.286	102489	96.6
39	274	1.8716	0.303	103706	97.7
39.5	276	1.9705	0.327	104332	98.3
40	277	2.0888	0.362	105067	99.0
40.5	278	2.2408	0.421	105372	99.3
41	280	2.4616	0.529	105761	99.6
41.5	282	2.8747	0.796	105926	99.8
42	285	3.7216	1.602	106134	100

## 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* (3<sup>rd</sup> 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* (4<sup>th</sup> 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* (3<sup>rd</sup> 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.