

2006
Grade Eight Proficiency
Assessment (GEPA)

TECHNICAL REPORT
March Administration



PTM#1506.20

Copyright © 2007 New Jersey
Department of Education

All rights reserved.

CHAPTER 1: INTRODUCTION	1
1.1 Description of the Grade Eight Proficiency Assessment (GEPA)	1
1.2 Purpose of the GEPA	4
1.3 GEPA Organizational Support.	5
CHAPTER 2: TEST DEVELOPMENT	6
2.1 Test Specifications.	6
Language Arts Literacy	7
Mathematics	8
Science	8
2.2 Development of Test Items	9
2.3 Item Review Process.	12
2.4 Operational Test Development	18
2.5 Review and Approve Operational Test Forms	20
2.6 Test Materials for Visually Impaired Students	20
CHAPTER 3: TEST ADMINISTRATION	22
3.1 Participation.	22
3.2 Test Security Procedures	23
3.3 Test Administration Procedures.	24
3.4 Test Accommodations	25
3.5 Results for Special Education Students and Section 504 Students Tested with Accommodations or Modifications	27
CHAPTER 4: SCORING	31
4.1 Multiple-choice Items	31
4.2 Open-ended Items	31
Scorer Selection	32
Rangefinding	32
Development of Scoring Guides	33
Team Leader Training and Qualifying	33
Scorer Training and Qualifying	34
Scoring Procedures and Paper Flow	35
Scorer Monitoring	36
Agreement Between Scorers for the Writing Tasks and Open-Ended Items	36
4.3 Quality Control Procedures in Data Preparation	36
Scanning and Scoring	38
NJDOE Quality Control of Score Reporting	39
CHAPTER 5: STANDARD SETTING	40
5.1 Overview of the Process	40
5.2 Procedures	40
Language Arts Literacy–Proficiency Level Descriptors	41
Mathematics–Proficiency Level Descriptors	42
Science–Proficiency Level Descriptors	43

<i>Judge Selection Process and Criteria</i>	44
<i>Holistic/Paper Sorting Methodology</i>	44
<i>Overview of the 8-Step Plan</i>	45
5.3 Results	47
CHAPTER 6: SCALING AND EQUATING	49
6.1 Scaling	49
6.2 Equating	49
<i>Mathematics and Science Equating Design</i>	50
<i>Language Arts Literacy Equating Design</i>	51
CHAPTER 7: TEST STATISTICS	53
7.1 Reliability of the Test Scores	53
CHAPTER 8: ITEM-LEVEL STATISTICS	56
8.1 Classical Item Statistics	56
8.2 Speededness	59
8.3 Intercorrelations	61
CHAPTER 9: TEST VALIDITY	63
9.1 Content and Curricular Validity (Evidence Based on Test Content)	65
9.2 Construct Validity (Evidence Based on Response and Evidence Based on Internal Structure)	66
9.3 Criterion-Related Validity (Evidence Based on Relations to Other Variables)	67
9.4 Consequential Validity Evidence (Evidence Based on Consequences of Testing) ..	70
CHAPTER 10: REPORTING	73
10.1 Information on the Reports	73
10.2 Types of Reports	75
<i>Cycle I Reports</i>	75
<i>Cycle II Reports</i>	80
Appendix A: Scoring Rubrics and 3rd Reader Score Calculation Charts ..	84
Appendix B: Cycle II Test Results	95
Appendix C: Modifications of Test Administration Procedures for Special Education Students and Students Eligible Under Section 504 of the Rehabilitation Act of 1973	115
Appendix D: Raw to Scale Scores Conversions	120
Appendix E: Scale Scores with Frequencies	127
Appendix F: Reporting Limited English Proficiency (LEP) and Special Education (SE)	134
References	137

Tables

TABLE 1.1	Number of Items and Approximate Times	1
TABLE 1.2	Total Student Group Testing in 2006	2
TABLE 2.1	Total Points Possible for the Language Arts Literacy Component of the GEPA	10
TABLE 2.2	Total Points Possible for the Mathematics Component of the GEPA	10
TABLE 2.3	Total Points Possible for the Science Component of the GEPA	10
TABLE 2.4	Item Development Goals for the 2006 Field Test	11
TABLE 2.5	District Factor Groups (DFG) Represented on the GEPA Content and Sensitivity Committees	12
TABLE 2.6	GEPA 2006 Content and Sensitivity Committee Meetings	13
TABLE 2.7	LANGUAGE ARTS LITERACY - READING – Number of Field Test Items Approved During Statistical Review	17
TABLE 2.8	LANGUAGE ARTS LITERACY - WRITING – Number of Field Test Items Approved During Statistical Review	17
TABLE 2.9	MATHEMATICS – Number of Field Test Items Approved During Statistical Review	19
TABLE 2.10	SCIENCE – Number of Field Test Items Approved During Statistical Review	19
TABLE 2.11	Operational Test Specifications	21
TABLE 3.1	Setting Accommodations and Modifications for Special Education Students and Section 504 Students	28
TABLE 3.2	Setting Accommodations for Special Education Students and Section 504 Students	28
TABLE 3.3	Scheduling Accommodations for Special Education Students and Section 504 Students	29
TABLE 3.4	Test Materials Modifications for Special Education Students and Section 504 Students	29
TABLE 3.5	Test Procedures Modifications for Special Education Students and Section 504 Students	30
TABLE 4.1	Number of Writing Prompts and Open-ended Items Scored	31
TABLE 4.2	Consistency Between Raters Scoring GEPA Writing Tasks and Open-Ended Items - March 2006	37
TABLE 5.1	Proficiency-Level Cut Scores	48
TABLE 5.2	Percentage of Students Achieving Each Performance Level	48
TABLE 6.1	Summary of Equating Statistics	52

TABLE 6.2 Cut Scores and Associated Thetas for Proficiency Levels	52
TABLE 7.1 Reliability Estimates and Standard Errors of Measurement (SEMs) for Content Areas and Clusters - 2006	54
TABLE 7.2 Reliability Estimates and Standard Errors of Measurement (SEMs) for Dichotomously Scored Items Within Content Clusters - 2006	55
TABLE 8.1 Item Difficulty and Discrimination Summary Statistics for Dichotomously Scored and Open-Ended Items by Test Section and Cluster - 2006	56
TABLE 8.2 Frequency Distributions of Item Difficulty and Item Discrimination by Content Cluster - 2006 Language Arts Literacy	57
TABLE 8.3 Frequency Distributions of Item Difficulty and Item Discrimination by Content Cluster - 2006 Mathematics	58
TABLE 8.4 Frequency Distributions of Item Difficulty and Item Discrimination by Content Cluster - 2006 Science	59
TABLE 8.5 Percentage of Students Omitting the Last Items of Each Test Part - 2006	60
TABLE 8.6 Intercorrelations Among Major Content Clusters and Item Types - 2006 . .	61
TABLE 8.7 Intercorrelations Among Content Areas and Clusters - 2006	62
TABLE 9.1 LANGUAGE ARTS LITERACY - Percentages of Students Across GEPA/HSPA Proficiency Levels.	68
TABLE 9.2 MATHEMATICS - Percentages of Students Across GEPA/HSPA Proficiency Levels.	69

Figures

FIGURE 2.1: GEPA Test Development Process.	7
FIGURE 2.2: Item Approval Before Field Test.	12
FIGURE 2.3: Item Approval Before Operational Test	16
FIGURE 9.1 LANGUAGE ARTS LITERACY - Longitudinal Graph by Economic Status . . .	71
FIGURE 9.2 MATHEMATICS - Longitudinal Graph by Economic Status.	72
FIGURE 9.3 SCIENCE - Longitudinal Graph by Economic Status	72
FIGURE 10.1: Individual Student Report (ISR Front).	76
FIGURE 10.2: Individual Student Report (ISR Back)	77
FIGURE 10.3: Student Roster (Mathematics)	79
FIGURE 10.4: Cluster Means Report	81
FIGURE 10.5: Performance by Demographic Groups.	83

CHAPTER 1: INTRODUCTION

1.1 Description of the Grade Eight Proficiency Assessment (GEPA)

The New Jersey Grade Eight Proficiency Assessment (GEPA) for the 2006 administration consisted of three content area tests – Language Arts Literacy, Mathematics, and Science. The GEPA is designed to provide an indication of the progress students are making in mastering the knowledge and skills described in New Jersey’s Core Curriculum Content Standards for these content areas.

The GEPA was administered between Monday, March 13 and Thursday, March 16, 2006, with make-up testing between Monday, March 20 and Thursday, March 23, 2006. March 2006 marked the eighth administration of the GEPA, which provides valuable information about student progress toward mastery of the skills required for high school graduation. Table 1.1 lists the number of test items and approximate testing times for the three content areas.

TABLE 1.1
Number of Items and Approximate Times

Content Areas	Items	Approximate Times and Days
Science	45 multiple-choice 3 open-ended Embedded field-test items	Monday morning 1 hour, 57 minutes
Mathematics	30 multiple-choice 6 open-ended Embedded field-test items	Tuesday morning 2 hours, 27 minutes
Language Arts Literacy	20 multiple-choice 4 open-ended 2 writing tasks Field-test component	Wednesday morning Thursday morning 2 hours, 12 minutes (per day)

The GEPA Language Arts Literacy measures both reading and writing. The Reading component requires students to read passages and to respond to related items. The passages are selected from published books, newspapers, and magazines, as well as everyday text. The Reading component includes both multiple-choice and open-ended items. The open-ended items require students to write a few sentences or a few paragraphs to answer a question about the text. The Writing component asks students to write two essays. All the tasks in the Writing component require students to write a response.

The GEPA Mathematics measures students’ abilities to solve problems using mathematical concepts. The components in this content area measure: Number and Numerical Operations; Geometry and Measurement; Patterns and Algebra; and Data Analysis, Probability, and Discrete Mathematics. Mathematics, like the Reading component of Language Arts Literacy, contains both multiple-choice and open-ended items. The open-ended items require students to solve a problem as well as explain their solution.

The GEPA Science measures students’ knowledge in Life Science, Physical Science, and Earth Science; and skill in Knowledge and Application. The Science content area contains both multiple-choice and open-ended items. The open-ended items require students to respond to a question as well as explain the answer.

Rubrics for scoring the GEPA open-ended items and writing prompts are included in Appendix A of this *Technical Report*.

Table 1.2 presents the statewide test results for the 2006 administration of the GEPA. This table shows the number and percentages of students in each of the Proficiency Levels – Partially Proficient, Proficient, and Advanced Proficient. The first column in Table 1.2 shows the total 109,091 enrolled students including 87,707 general education students, 18,327 special education students, and 3,188 limited English students. “General Education” excludes students coded as special education (SE) or limited English proficient (LEP) on their answer folders. “Special Education” includes students coded as SE. “Limited English Proficient” includes students coded as LEP. “Total Students” refers to all students tested (general education, special education, and current LEP students).

Following the Number Enrolled column are the columns for Number of APA Students, Number Not Present, and Number of Voids. Number enrolled represents total number of answer folders returned. The number of APA (Alternate Proficiency Assessment) students shows the number of answer folders marked for students taking the APA rather than GEPA for each content area.

TABLE 1.2
Total Student Group Testing in 2006

TESTS	NUMBER OF STUDENTS ENROLLED	NUMBER OF APA STUDENTS	NUMBER NOT PRESENT	NUMBER OF VOIDS	NUMBER OF VALID SCALE SCORES	PROFICIENCY LEVELS						MEAN SCALE SCORE
						PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		
						NO.	%	NO.	%	NO.	%	
LANGUAGE ARTS LITERACY												
General Education	87,707	0	345	313	87,049	13,936	16.0	64,122	73.7	8,991	10.3	221.4
Special Education	18,327	689	216	346	17,076	11,491	67.3	5,493	32.2	92	0.5	184.2
LEP Current & Former	4,007	11	45	708	3,243	2,468	76.1	759	23.4	16	0.5	177.2
LEP Current	3,188	7	43	701	2,437	2,052	84.2	379	15.6	6	0.2	170.7
LEP Former	819	4	2	7	806	416	51.6	380	47.1	10	1.2	196.7
Total Students^a	109,091	689	604	1,351	106,447	27,375	25.7	69,983	65.7	9,089	8.5	214.3
MATHEMATICS												
General Education	87,707	0	416	66	87,225	22,929	26.3	42,510	48.7	21,786	25.0	222.1
Special Education	18,327	697	244	107	17,279	12,955	75.0	3,791	21.9	533	3.1	182.9
LEP Current & Former	4,007	11	27	9	3,960	2,868	72.4	862	21.8	230	5.8	185.0
LEP Current	3,188	7	26	6	3,149	2,436	77.4	566	18.0	147	4.7	180.8
LEP Former	819	4	1	3	811	432	53.3	296	36.5	83	10.2	201.1
Total Students^b	109,091	697	685	179	107,530	38,220	35.5	46,845	43.6	22,465	20.9	214.6
SCIENCE												
General Education	87,707	0	414	58	87,235	11,709	13.4	54,391	62.3	21,135	24.2	228.9
Special Education	18,327	665	253	76	17,333	8,702	50.2	7,787	44.9	844	4.9	200.6
LEP Current & Former	4,007	10	24	3	3,970	2,644	66.6	1,228	30.9	98	2.5	191.3
LEP Current	3,188	6	24	2	3,156	2,327	73.7	785	24.9	44	1.4	187.2
LEP Former	819	4	0	1	814	317	38.9	443	54.4	54	6.6	206.9
Total Students^c	109,091	665	690	136	107,600	22,638	21.0	62,939	58.5	22,023	20.5	223.2

^a The number of Valid Scale Scores includes 115 students who are both Special Education and Limited English Proficient.

^b The number of Valid Scale Scores includes 123 students who are both Special Education and Limited English Proficient.

^c The number of Valid Scale Scores includes 124 students who are both Special Education and Limited English Proficient.

Number not present indicates the number of answer folders returned that were totally blank excluding answer folders coded as APA. A student's answer folder can be voided at the time of testing due to illness, cheating or disruptive behavior, or some other reason. If a student's answer folder is voided, no total test score for that student is reported for the content area. A void code is printed in place of the total test score on the student's individual reports.

During the scoring process, a void code is given if a student's answer folder showed less than 20 percent of the items were attempted on the Mathematics or Science content area tests. During the 2006 administration, 10 Mathematics and 13 Science tests were voided due to the attempted criteria.

For Language Arts Literacy, if a student attempted less than 20 percent of the items on one or two testing days but attempted 20 percent or more on the other testing day, a Void code appeared instead of a total test score on the student's reports. However, cluster scores are provided for parts of the Language Arts Literacy that are attempted. During the 2006 administration, 176 Language Arts Literacy tests were voided due to the attempted criteria for Day 1 and 262 Language Arts Literacy tests were voided due to the attempted criteria for Day 2.

Table 1.2 shows that a total of 106,447 students had valid scale scores in Language Arts Literacy, 107,530 students had valid scale scores in Mathematics, and 107,600 students had valid scale scores in Science. The number of valid scale scores is the number enrolled excluding the number of APA students, number not present, and number of voids.

Performance data shown in the Proficiency Levels columns include students who received valid scale scores. The number of students who scored in each proficiency level excludes students coded as APA. Because each content area is independent, students may receive a scale score in one content area, but not in others.

The total GEPA Language Arts Literacy, Mathematics, and Science scores are reported as scale scores with a range of 100 to 300. Scale scores of 100 and 300 are a theoretical floor and ceiling, which may not actually be observed. The score ranges are as follows:

<i>Advanced Proficient</i>	<i>250–300</i>
<i>Proficient</i>	<i>200–249</i>
<i>Partially Proficient</i>	<i>100–199</i>

A series of tables summarizing the test results for the State (general education students, special education students, limited English proficient students, and total students), District Factor Groups, Special Needs Districts, and All Other (Non Special Needs) Districts appears in Appendix B. See <http://www.state.nj.us/njded/finance/> for information about District Factor Groups and Special Needs Districts (Abbott Districts).

Note that the percentages shown in tables throughout this Technical Report may not total to 100 due to rounding.

1.2 Purpose of the GEPA

The GEPA serves as a primary indicator for identifying those students who may need instructional intervention in the three content areas of Language Arts Literacy, Mathematics, and Science. The test also serves as an indicator for determining which local education programs may require revisions to ensure that instructional programs are aligned with the Core Curriculum Content Standards. The GEPA is designed to evaluate the progress students are making in mastering the knowledge and skills required by the end of eighth grade. Also, the GEPA provides an indication of students' progress in the skills required to pass the High School Proficiency Assessment.

Three proficiency levels have been determined for each of the content areas of the GEPA: Partially Proficient, Proficient, and Advanced Proficient. Students scoring in the lowest level, Partially Proficient, are considered below the state minimum level of proficiency. These students may need instructional intervention. Instructional decisions for all students are determined only after additional information is considered, e.g., classroom tests, teacher observations.

In 1996, the State Board of Education adopted Core Curriculum Content Standards to describe what all students should know and be able to do at the end of fourth grade, eighth grade, and upon completion of a New Jersey public school education. The Core Curriculum Standards delineate New Jersey's expectations for student learning. All New Jersey school districts are required to organize instruction and design curricula so that virtually all students achieve these content standards. The Core Curriculum Content Standards defined the development of three statewide assessments: the Elementary School Proficiency Assessment Program, which was administered from 1997-2002; the GEPA, which replaced the Early Warning Test (EWT) in 1998; and the High School Proficiency Assessment, which replaced the High School Proficiency Test as the state's graduation requirement for all students who entered the eleventh grade in the fall of 2001.

Previously, in 1988, the New Jersey Legislature passed a law that established the Early Warning Test. The Legislature moved the High School Proficiency Test from the ninth grade to the eleventh grade. The Grade 11 High School Proficiency Test assessed essential reading, mathematics, and writing skills. It served as a graduation requirement for all public school students in New Jersey who entered ninth grade on or after September 1, 1991, and prior to the fall of 2001.

The Early Warning Test was similar to the High School Proficiency Test in eleventh grade because it also measured basic skills in reading, mathematics, and writing. The Early Warning Test was administered to all eighth-grade students each spring to determine whether they were making satisfactory progress in mastering the skills they would need to pass the High School Proficiency Test in the eleventh grade. The Early Warning Test was first administered as an operational test in March 1994.

Following the adoption of the Core Curriculum Standards in 1996, the development of the GEPA was defined. The GEPA was initially administered as field tests in Language Arts Literacy and Mathematics. In March 1999, the GEPA was administered for the first time as an operational assessment. Additional field tests in Language Arts Literacy, Mathematics, and Science were also administered and the GEPA Speaking assessment was pilot tested. In March 2000, Science was included in GEPA as an operational test for the first time.

Because the State Board required that the Core Curriculum Content Standards be reviewed and revised every five years, a review process began in May 2001 involving teachers, school administrators, students, parents, and representatives from business, higher education, and the community.

The language arts literacy, mathematics, and science standards were adopted by the State Board of Education in July 2002. In April 2004, the language arts literacy standards were revised to comply with the requirements of the No Child Left Behind Act of 2001 (NCLB) and readopted by the Board.

The GEPA administration in 2006 included field test items that were aligned with the new Core Curriculum Content Standards for language arts literacy, mathematics, and science. The GEPA test development procedures are detailed in Chapter 2 of this *Technical Report*.

1.3 GEPA Organizational Support

New Jersey Department of Education (NJDOE) The GEPA 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 GEPA activities, the staff is extensively involved in numerous test review, security, and quality control procedures.

Pearson Educational Measurement (PEM—previously NCS Pearson) Pearson Educational Measurement is the primary contractor working in partnership with Measurement Incorporated (MI) and Assessment and Evaluation Services (AES). In 1998, the contract for developing and administering the GEPA was awarded to NCS Pearson which became Pearson Educational Measurement in 2003. Major Pearson Educational Measurement activities include the following:

- Supporting and monitoring the test development cycle and subcontractor efforts toward content development
- Printing test books and ancillary materials required for the GEPA
- Distributing assessment materials in a secure manner and in appropriate amounts based on the district quantity survey results
- Supporting the regional workshops that inform district test coordinators about the GEPA program
- Receiving, scanning, editing, and scoring the answer documents using clearly defined quality control procedures
- Packaging and transporting open-ended responses to be hand-scored
- Providing accurate reports of test results to New Jersey pupils, parents/guardians, schools, districts, and the state

Measurement Incorporated (MI) MI provides item development and scores all open-ended responses for the GEPA program. Items developed include multiple-choice and constructed-response items for Language Arts Literacy, Mathematics, and Science; and writing prompts for Language Arts Literacy. MI scoring directors, NJDOE Office of Evaluation and Assessment content specialists, and New Jersey teachers use rangefinding procedures to prepare for scoring the GEPA open-ended items.

Assessment and Evaluation Services (AES) AES is responsible for GEPA technical activities such as specifying the item selection for the operational tests, equating the test forms, and developing the scale score conversion tables.

CHAPTER 2: TEST DEVELOPMENT

The New Jersey Department of Education has developed a comprehensive set of assessments that measure student achievement of the Core Curriculum Content Standards. The validity of the GEPA is therefore based on the alignment of the GEPA, the Core Curriculum Content Standards, and the knowledge and skills expected of eighth-grade students.

This chapter 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 assisted 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.

2.1 Test Specifications

The GEPA content areas of Language Arts Literacy, Mathematics, and Science were designed from their inception in 1997 to align with the original Core Curriculum Content Standards adopted by the New Jersey State Board of Education in 1996. The State Board required that the Core Curriculum Content Standards be reviewed every five years. New standards for the three content areas were adopted by the Board in July 2002. To comply with requirements of the federal No Child Left Behind Act of 2001 (NCLB), the Language Arts Literacy standards were also revised in April 2004.

The Core Curriculum Content Standards were developed by teachers and other educational professionals from New Jersey. The Core Curriculum Content Standards outline what students should know and be able to do at a certain grade level. The questions on the GEPA can contain items/concepts included in the grade eight standards as well as for those standards listed in the prior grade standards.

The GEPA was first administered as an operational assessment in 1999. Prior to that time, Language Arts Literacy and Mathematics was administered to all eighth-grade students as field tests and “due-notice” administrations. Science was initially field tested in 1999. The purpose of due-notice administrations was to help school districts identify potential gaps between their curriculum and the test objectives, and to allow schools time to modify their curriculum and instructional practices to meet the needs of students before the first operational assessment. Field test items for Language Arts Literacy, Mathematics, and Science continued to be included with the GEPA 2000 – 2006 test administrations.

Following adoption of the original Core Curriculum Content Standards in 1996, the New Jersey Assessment Content Committees met through 1997 to develop 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, item formats, and test item scoring. 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 eighth-grade students and is composed of test questions that are age- and grade-appropriate. The material in the three directories of test specifications and sample items is designed for use by curriculum specialists and teachers to improve instruction at the district, school, and classroom levels. Figure 2.1 summarizes the steps of the test development process beginning with the development of the Core Curriculum Content Standards and ending with an operational GEPA test form. Brief descriptions of the test content measured in Language Arts Literacy, Mathematics, and Science are presented in the following sections.

Language Arts Literacy

Language Arts Literacy measures students' achievements in reading and writing. Language Arts Literacy currently assesses knowledge and skills in two content clusters:

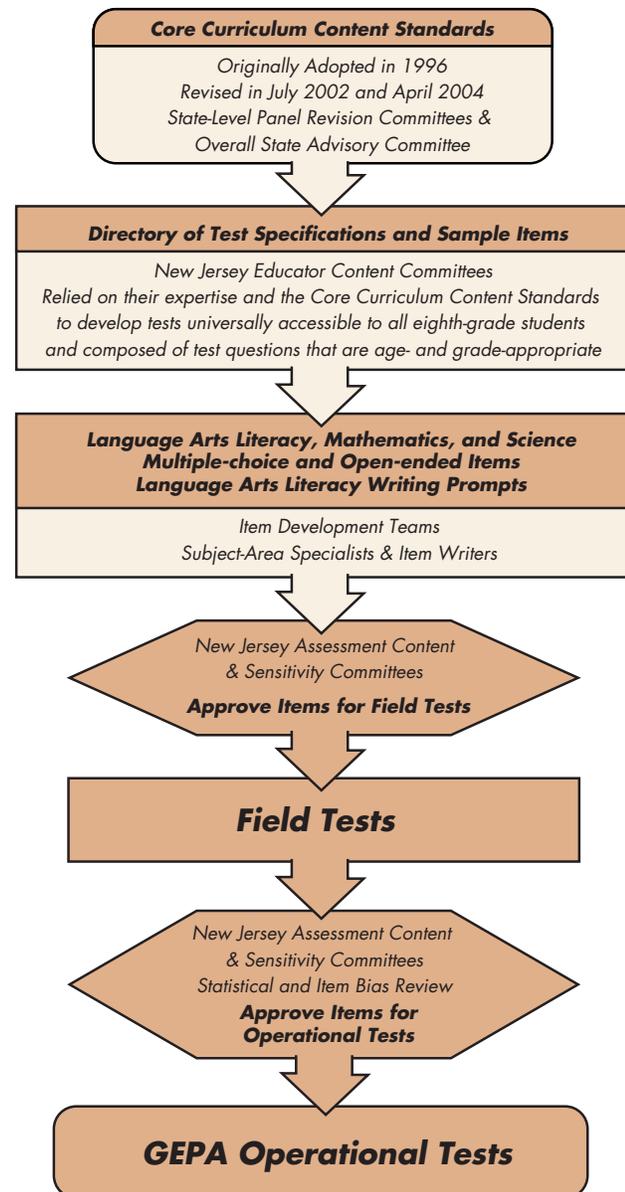
- Reading
- Writing

The Reading cluster consists of a narrative reading passage with ten multiple-choice and two open-ended items, and a persuasive reading passage with ten multiple-choice and two open-ended items. The passages are selected from published sources such as books, newspapers, magazines, and the Internet.

The Writing cluster for GEPA consists of two writing activities: a writing/persuade task in response to a prompt and a writing/speculate task in response to a picture.

For an in-depth description of the Language Arts Literacy assessment, refer to the *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* (February 1998). The directory is available online at <http://www.njpep.org/assessment/TestSpecs/LangArts/TOC.html>, or by calling the New Jersey Department of Education, Publications Office, (609) 984-0549.

FIGURE 2.1
GEPA Test Development Process



The Technical Advisory Committee (TAC) advises and assists the Office of Evaluation and Assessment in the development and implementation of the statewide testing program. TAC reviews and provides suggestions for each of the stages listed in the GEPA Test Development Process.

Mathematics

Mathematics measures students' ability to solve problems by applying mathematical concepts. The GEPA Mathematics assessment measures knowledge and skills in four content clusters:

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

Mathematics items are also classified and reported as Knowledge (requiring conceptual understanding or procedural knowledge) and Problem Solving (applying mathematical concepts). For the operational test, there are a total of 30 multiple-choice and 6 open-ended items in Mathematics.

For an in-depth description of the GEPA Mathematics assessment, refer to the *Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and the High School Proficiency Assessment (HSPA) in Mathematics* (February 1998). The directory is available online at <http://www.njpep.org/assessment/TestSpecs/MathTestSpec/GEPAMath/MathIndex.html>, or by calling the New Jersey Department of Education, Publications Office, (609) 984-0549.

Additional information about the GEPA test specifications is included at http://www.njpep.org/assessment/TestSpecs/MathTestSpec/GEPAMath_sample_questions/worddocs/GEPA%20Math%202005%20presentation.ppt

Science

Science measures knowledge and skills in three content clusters:

- Life Science
- Physical Science
- Earth Science

Science items are also classified and reported as Knowledge (Comprehension and Science, Society/Technology) and Application (Habits of Mind/Inquiry and Mathematics). For the operational test, there are a total of 45 multiple-choice and 3 open-ended items in Science.

For an in-depth description of the Science assessment, refer to the *Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and the High School Proficiency Assessment (HSPA) in Science* (February 1998). The directory is available online at http://www.njpep.org/assessment/TestSpecs/science_test_specs/Science_GEPA_HSPA/, or by calling the New Jersey Department of Education, Publications Office, (609) 984-0549.

Additional information about the GEPA test specifications is included at <http://www.njpep.org/assessment/TestSpecs/ScienceGEPA/index.html>

Tables 2.1, 2.2, and 2.3 summarize the total points possible for Language Arts Literacy, Mathematics, and Science of the content areas of the operational GEPA administered in 2006.

2.2 Development of Test Items

The 2006 GEPA consists of two types of items:

- operational test items used to determine students' scores and
- field test items evaluated for use as future operational test items.

The 2006 operational test for Language Arts Literacy, Mathematics, and Science was composed of items field tested through 2005. The item development teams consisted of subject-area specialists and consulting item writers. These writers were teachers or former teachers with a great deal of specialized knowledge (e.g., education and training, years of classroom experience, familiarity with the student population, knowledge of the content area, and understanding of the pedagogy that defines the discipline) concerning their area of content expertise.

Each of the content areas consists of multiple-choice and open-ended items. The multiple-choice items are designed to measure those skills determined to be best measured by such item types, and the open-ended items are developed to measure those skills requiring students to do more than select a correct answer. That is, the open-ended items are designed to tap more complex and integrated skills. Language Arts Literacy includes a writing/persuade task and a writing/speculate task in response to a picture.

The Measurement Incorporated/Pearson Educational Measurement item development process for each testing cycle begins with a formal review of the Core Curriculum Content Standards and the three directories of test specifications. Item-writing training sessions typically last from 8 to 16 hours over two days. The respective test development specialist for each content area conducts the training session. Between the first and second sessions, preliminary versions of test items developed in the first session are evaluated. At the second session, the training is focused on the items developed in the first session.

At the training, each consulting item writer is asked to sign a Letter of Agreement. This letter specifies the confidentiality and security regulations. The agreement also outlines the ownership regulations. No confidential materials related to the project are released without explicit approval of the NJDOE Office of Evaluation and Assessment.

During the training, each item writer is given the following information:

- An overview of the GEPA
- Final test blueprint for each subject-area test and item specifications
- A description of the item formats used, including important characteristics of each format
- A description of the item writing process and measures to take to avoid writing biased items
- A listing of the security procedures followed during the item development process.

Important guidelines for the GEPA item development and test structure are outlined below.

1. Items are written to reflect what students know and understand based on classroom instruction and their mastery of skills included in the Core Curriculum Content Standards. Items are also designed to assess higher-order or critical thinking skills in varied contexts that students are likely to understand; yet, they are based upon solid theoretical frameworks.

TABLE 2.1

Total Points Possible for the Language Arts Literacy Component of the GEPA

Language Arts Literacy		
Total	54 points	
Reading	36 points*	
Writing	18 points*	
Writing/Speculate	6 points*	1 – 6 points, ratings averaged
Writing/Persuade	12 points*	1 – 6 points, ratings summed

Interpreting Text	20 points*	
Analyzing/Critiquing Text	16 points*	

* 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 GEPA can contribute to a cluster above the line (for example, Reading) as well as a cluster below the line (for example, Interpreting Text), each item is counted only once in the total score.

TABLE 2.2

Total Points Possible for the Mathematics Component of the GEPA

Mathematics		
Total	48 points	
Number and Numerical Operations	12 points*	
Geometry and Measurement	12 points*	
Patterns and Algebra	12 points*	
Data Analysis, Probability, and Discrete Mathematics	12 points*	

Knowledge	48 points*	
Problem Solving	36 points*	

* 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). All Mathematics items are classified as Knowledge because all items require conceptual understanding or procedural knowledge. Some items also measure Problem Solving. Each Mathematics item counts only once in the total score.

TABLE 2.3

Total Points Possible for the Science Component of the GEPA

Science		
Total	54 points	
Life	22 points*	
Physical	16 points*	
Earth	16 points*	

Knowledge	12 points*	
Application	42 points*	

* 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 GEPA can contribute to a cluster above the line (for example, Life) as well as a cluster below the dotted line (for example, Knowledge), each item is counted only once in the total score.

2. For each content area, the multiple-choice items represent a range of difficulty. For example, approximately 25 percent of the items are relatively easy, 50 percent of the items are somewhat difficult, and 25 percent of the items are difficult. This range of difficulty provides for a distribution of items with p-values from approximately 0.30 to 0.95. This distribution allows for a range of difficulty that supports the established proficiency levels, yet is not so difficult that low-achieving students cannot be assessed adequately.
3. Item content for all of the items, including the writing-task prompts, is carefully reviewed to ensure that the items are free from gender, ethnic and regional bias. Across all content areas of the GEPA and in any test material presented, there is a balance of gender and active/passive roles by gender.
4. Measurement Incorporated/Pearson Educational Measurement construct initial rubrics for each open-ended item in Language Arts Literacy, Mathematics, and Science.
5. Writing task prompts for Language Arts Literacy are written in such a way that they focus on experiences that eighth-grade students may have every day. However, care must be taken to ensure that the writing task prompts are not intrusive in nature and do not elicit personal information of a biographical, religious, political, or affective nature. Topics must be chosen so that no group of eighth-grade students is put at a subject-related disadvantage. Instead, each writing task prompt is designed to sample the skills and abilities demanded of eighth-grade students. Each writing task is developmentally appropriate for students in both the academic and nonacademic environments.

As items are developed, Measurement Incorporated/ Pearson Educational Measurement document each item's relevancy to the Core Curriculum Content Standards and to the directories of test specifications. During this process, each item is assigned a unique item ID number or coding system number. This unique number identifies the following: content area, skill measured, standard, and associated materials such as a reading passage, artwork, or display of data. The number is used to track the item throughout the development process and its eventual use on the operational test.

All items prepared by item writers are reviewed, revised, and edited by the subject area specialists and editors prior to review by the New Jersey Assessment Content Review Committees. Also, the New Jersey Assessment Sensitivity Review Committee approves passages used on the Language Arts Literacy section.

In preparation for the 2006 field test items, a total of 154 Language Arts Literacy, 120 Mathematics, and 160 Science items were requested by the NJDOE Office of Evaluation and Assessment staff. The request for the Writing component of Language Arts Literacy included five prompts for the writing/persuade task and five pictures for the writing/speculate task. Table 2.4 shows the number of multiple-choice and open-ended items specified for each content area.

TABLE 2.4

Item Development Goals for the 2006 Field Test

	<i>Multiple-choice Items</i>	<i>Open-ended Items</i>	<i>Total Items</i>
	<i>Goal</i>	<i>Goal</i>	<i>Goal</i>
<i>Language Arts Literacy</i>	120	34	154
<i>Mathematics</i>	100	20	120
<i>Science</i>	150	10	160
TOTAL	370	64	434

2.3 Item Review Process

The New Jersey Assessment Content Committee members provide expert judgments on the alignment of each test item with the Core Curriculum Content Standards and the content-specific test specifications. The committee members represent school districts across all District Factor Groups. Table 2.5 shows the District Factor Groups represented on each of the Content and Sensitivity Committees.

TABLE 2.5

District Factor Groups (DFG) Represented on the GEPA Content and Sensitivity Committees

DFG	Language Arts Literacy	Mathematics	Science	Sensitivity	Total
A	1	1	4	1	7
B	3	2	2	2	9
CD	1	1	1	0	3
DE	2	2	0	0	4
FG	3	3	1	1	8
GH	2	3	1	0	6
I	2	0	2	1	5
J	0	1	0	0	1
Retirees	4	5	3	7	19
Private School	0	0	1	0	1
Not in Districts	0	2	1	1	4
Total	18	20	16	13	67

Committee members sign a Confidentiality and Security Agreement noting they must maintain the security of the testing materials by not discussing and disclosing any confidential information related to the program.

FIGURE 2.2

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

Prior to field testing, all items are reviewed by the Office of Evaluation and Assessment staff and committee members. Each test item is reviewed 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 2.2 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 in a field test. The percentage of items accepted for field testing depends on the content area and the item type. The range of acceptance generally is 60-80% at this item review stage. During review, committee members approve items, amend or revise items, or reject items.

Items field tested during March 2006 were reviewed by the committees during spring and summer 2005. The committees met in August 2006 to review item statistics from the March 2006 field testing. The statistical item review meetings are listed in Table 2.6. Because the Office of Evaluation and Assessment requested no new item development for 2007 field testing, no item development meetings were held during spring and summer 2006.

At the statistical review, committee members consider how well students did on each field test question in comparison to the other questions on the GEPA. If an item yields good statistics, it will become part of the operational pool for future GEPA tests. Otherwise, it will be eliminated or revised and re-field tested.

Prior to field test statistical review, the field-tested open-ended items and writing prompts must go through rangefinding to determine the scores on sample student responses. The field test rangefinding process involves scoring 30 student responses for each of the open-ended items and writing prompts. These 30 responses are selected to represent the wide range of responses to that item. The papers are scored by one or two content committee members, the NJDOE Content Coordinator, and representatives from Measurement Incorporated.

TABLE 2.6
GEPA 2006 Content and Sensitivity Committee Meetings

Language Arts Literacy Committee
Statistical Item Review Tue – Wed, August 8 – 9
Mathematics Committee
Statistical Item Review Tue, August 8
Science Committee
Statistical Item Review Tue – Wed, August 8 – 9
Sensitivity Committee
Statistical Item Review Tue – Wed, August 8 – 9

Because the Office of Evaluation and Assessment requested no new item development for 2007 field testing, only statistical item review meetings were held during spring and summer 2006.

In Language Arts Literacy, the responses are scored according to the generic rubric for either reading or writing as appropriate. Use of these generic rubrics ensures that student responses are scored in the same way for the demonstration of the same level of knowledge and skills regardless of the prompt or the year.

For Mathematics and Science, each item has a unique scoring rubric, based on the generic one for each area. During rangefinding, the item specific rubric is refined, if necessary, to define each score point clearly. The rangefinding process aids in delineating between a 0 & 1, 1 & 2, and a 2 & 3 score point response. The holistic scoring guide is used quite often to refine the tenuous line between the score points.

For all content areas, the scored field test responses and the rubrics are used to create the holistic scoring guide, which is used to help refine the lines between the score points. This guide is then used to train the scorers of that item. If there is any problem or question with the scoring of a student's response, the NJDOE Content Coordinator is contacted and makes a final decision for the score of that paper. After the open-ended papers have been scored, the scorers discuss the types of responses and problems, if any, found during scoring of each item. The scoring director then writes a brief summary of these comments and sends it, along with a copy of each item, rubric, sample answer, and rangefinding paper to the statistics review. Other than this packet, the same field test review procedures are used for the open-ended and multiple-choice items.

Pearson Educational Measurement computes item means, response frequencies, biserial correlations, and other descriptive statistics. Prior to the presentation of items and statistics to reviewers, the NJDOE Office of Evaluation and Assessment defined boundaries within which item statistics should fall. In general, items with p-values below 0.30 or above 0.95 were considered 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 operational test total scores.

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 would lead us to expect. This statistic takes into consideration both group membership (by race or by gender) and ability. 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 are somewhat suspect. 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. The Mantel-Haenszel statistic is used at Educational Testing Service (ETS) as a classification determination of category A, B, and C as described by Zieky (1993):

-
- Category A) MH D-DIF not significantly different from zero
OR
absolute value less than 1.0*
- Category B) MH D-DIF significantly different from zero and absolute value of at least 1.0
AND EITHER
(1) less than 1.5
OR
(2) not significantly greater than 1.0*
- Category C) MH D-DIF significantly greater than 1.0
AND
absolute value 1.5 or more. (p. 342)*
-

For every open-ended item and writing prompt, the Sensitivity Committee reviews frequency distributions for the range of scores of the following student groups: total, white, African American, Hispanic, Asian, American Indian, male, and female.

For the multiple-choice items field tested during 2006, fourteen items in Language Arts Literacy, four items in Mathematics, and five items in Science were flagged. The Sensitivity Committee marked “Do Not Use” or “Revise and Re-field Test” for three of the flagged Language Arts Literacy items. The Language Arts Literacy Committee marked “Do Not Use” for these three items as well as four more of the fourteen flagged items. Both the Sensitivity and Mathematics Committees marked “Do Not Use” for one of the four flagged Mathematics items. The Sensitivity and Mathematics Committees approved the other three flagged Mathematics items. The Sensitivity and Science Committees approved the five flagged Science items.

FIGURE 2.3

Item Approval Before Operational Test

Sensitivity		Content	
*Comments		*Comments	
Sensitivity Issue <input type="checkbox"/> Yes <input type="checkbox"/> No		Appropriate Difficulty <input type="checkbox"/> Yes <input type="checkbox"/> No	
If yes, identify category and explain*		PVal = Biserial =	
Mantel-Haenszel Category C <input type="checkbox"/> W-AA <input type="checkbox"/> W-H <input type="checkbox"/> M-F			
<input type="checkbox"/> Yes <input type="checkbox"/> No	Definitely Use		<input type="checkbox"/> Yes <input type="checkbox"/> No
<input type="checkbox"/> Yes <input type="checkbox"/> No	Revise and Use With Approval**		<input type="checkbox"/> Yes <input type="checkbox"/> No
<input type="checkbox"/> Yes <input type="checkbox"/> No	Revise and Re-Field Test		<input type="checkbox"/> Yes <input type="checkbox"/> No
<input type="checkbox"/> Yes <input type="checkbox"/> No	Do Not Use*		<input type="checkbox"/> Yes <input type="checkbox"/> No

Sensitivity Sign-off

Date

Content Chairperson's Signature

Date

**Requires director's approval

Figure 2.3 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.

Tables 2.7 – 2.10 present the number of items field tested during the administration.

Table 2.7 shows 140 multiple-choice items and 28 open-ended items were field tested for the Reading component of Language Arts Literacy, which included three narrative passages and four persuasive passages. During the statistical review, the Language Arts Literacy committee approved all narrative passages and three persuasive passages for operational tests. Two open-ended items for the fourth persuasive passage were dropped due to a poor range of students’ scores during field test rangefinding. During the statistical review, the Language Arts Literacy committee determined that one of the two companion pieces for the fourth passage lacked a strong point of view. Therefore, the committee concluded that the fourth persuasive passage with its items could not be approved for operational use.

Table 2.8 shows the results of the Writing component of Language Arts Literacy from the field tested two pictures for the writing/speculate task and four prompts for the writing/persuade task. All speculative (picture) prompts and persuasive prompts were approved for operational tests.

TABLE 2.7

LANGUAGE ARTS LITERACY - READING
Number of Field Test Items Approved During Statistical Review

Passages	Field-Tested		Approved		Revise & Re-Field Test		Do Not Use	
	MC	OE	MC	OE	MC	OE	MC	OE
Narrative 1	20	4	17	2	0	0	3	2
Narrative 2	20	4	19	3	0	0	1	1
Narrative 3	20	4	16	3	0	0	4	1
Persuasive 1	20	4	17	4	0	0	3	0
Persuasive 2	20	4	18	3	0	0	2	1
Persuasive 3	20	4	16	3	0	0	4	1
Persuasive 4	20	4	0	0	0	0	20	4
TOTAL	140	28	103	18	0	0	37	10

TABLE 2.8

LANGUAGE ARTS LITERACY - WRITING
Number of Field Test Items Approved During Statistical Review

Prompts	Field Tested	Approved	Revise & Re-field Test	Do Not Use
Speculate (Picture)	2	2	0	0
Persuade	4	4	0	0
TOTAL	6	6	0	0

Table 2.9 reports the results by content cluster for the 110 multiple-choice items and 21 open-ended Mathematics items field tested in 2006. Each content cluster is further divided into strands. Information about the new test specifications, including the associated strands, is located at <http://www.njpep.org/assessment/TestSpecs/MathTestSpec/GEPAMath/Macros.html>. Table 2.9 indicates that 80% Mathematics multiple-choice items and 19% Mathematics open-ended items were approved for an operational base test.

Table 2.10 shows that 180 multiple-choice and 11 open-ended Science items were field tested in 2006. This indicates that 75% Science multiple-choice items and 27.3% Science open-ended items were approved for an operational test. The number of Science items field tested for each content cluster as well as by knowledge skill and application skill is shown in the table.

Information about the science test specifications is located at http://www.njpep.org/assessment/TestSpecs/ScienceGEPA/TestSpecsRev9_04.doc

2.4 Operational Test Development

Following the 1998 through 2001 administrations, GEPA examiners completed a feedback form seeking suggestions and concerns related to the testing procedures. Questions related to timing, directions, and answer documents were asked specifically for each content area tested. Also, examiners were asked to identify questions that arose on issues and topics not addressed in the test booklets, directions, or coordinator or examiner manuals.

A sample of the 2001 questions is provided below:

- Was the time allotted for students to complete the test sufficient?
 - too much time
 - time about right
 - too little time
- Were the directions clear?
 - yes, directions were clear
 - no, directions were somewhat confusing
- Was the space provided for student responses in the answer folder sufficient?
 - adequate space
 - not enough space

Information from the examiners' responses assisted the Office of Evaluation and Assessment with determining the operational testing procedures.

TABLE 2.9
MATHEMATICS
Number of Field Test Items Approved During Statistical Review

Content Cluster	Strand	Field-Tested		Approved		Revise & Re-Field Test		Do Not Use	
		MC	OE	MC	OE	MC	OE	MC	OE
Number and Numerical Operations	A	0	2	0	0	0	0	0	2
	B	0	1	0	0	0	0	0	1
	C	0	2	0	0	0	1	0	1
Geometry and Measurement	A	0	1	0	1	0	0	0	0
	B	0	2	0	0	0	1	0	1
	C	32	2	28	0	0	1	4	1
	D	27	3	20	0	0	0	7	3
	E	5	1	4	0	0	0	1	1
Patterns and Algebra	A	1	0	1	0	0	0	0	0
	B	20	4	16	1	0	1	4	2
	C	0	1	0	1	0	0	0	0
	D	6	0	3	0	0	0	3	0
Data Analysis, Probability, and Discrete Mathematics	A	0	0	0	0	0	0	0	0
	B	0	0	0	0	0	0	0	0
	C	0	0	0	0	0	0	0	0
	D	19	2	16	1	0	0	3	1
TOTAL		110	21	88	4	0	4	22	13

TABLE 2.10
SCIENCE
Number of Field Test Items Approved During Statistical Review

Content and Skill Clusters		Field-Tested		Approved		Revise & Re-Field Test		Do Not Use	
		MC	OE	MC	OE	MC	OE	MC	OE
Life									
	Knowledge	18	0	11	0	2	0	5	0
	Application	79	5	62	2	10	3	7	0
Physical									
	Knowledge	12	0	8	0	2	0	2	0
	Application	47	5	38	1	7	1	2	3
Earth									
	Knowledge	4	0	2	0	1	0	1	0
	Application	20	1	14	0	3	1	3	0
TOTAL		180	11	135	3	25	5	20	3

The GEPA Content Committees assisted with recommending the emphases and priorities reflected in the number of items for each item type and cluster on the operational test. The operational test specifications appear in Table 2.11.

Following the approval of test items for use on operational tests by the Content and Sensitivity Review Committees, Assessment and Evaluation Services (AES) selected items for each GEPA administration to meet test specifications for Language Arts Literacy, Mathematics, and Science.

Relevant considerations for operational test development included content quality and scope, cluster representation, and appropriate item difficulty indices. The new operational test was parallel to the content, format, and statistical characteristics of the previous operational forms. Selecting test items for the operational tests is an iterative process to create test forms that are the perfect combination of content and statistical information. Through the iterative process, item content took precedence over statistical characteristics.

The operational test development used the Rasch model to pre-equate cluster and total test scores. Rasch item difficulty statistics were calibrated to the previous test administration. Common items were chosen to link the Mathematics and Science operational tests to previous forms for equating purposes. For Language Arts Literacy, the forward and backward items for equating purposes were specified. For each operational test, AES produces a spreadsheet that includes the following information for both the previous operational test and newly developed operational test.

- Item identifier with item type (multiple-choice or open-ended), content clusters, and skill clusters
- Common items for equating
- P-values and biserial correlations
- Item difficulties with sums and averages for clusters and total test

2.5 Review and Approve Operational Test Forms

The Office of Evaluation and Assessment approved the operational test forms for each GEPA administration. AES and PEM assisted with quality control that included:

- Confirm that each test item appears on the operational test as it was approved by the Content and Sensitivity Review Committees.
- Confirm that all test specification requirements are met.
- Check adequacy of common item set (i.e., in terms of size, content and skill representation)
- Double-check that the item and mean difficulty levels are accurate and meet requirements.
- Take the test to be certain all content considerations including content/skill/topic balance, correct keys, no clueing, and correct graphics are met.

2.6 Test Materials for Visually Impaired Students

The Office of Evaluation and Assessment works with the New Jersey Commission for the Blind and Visually Impaired to identify items with graphs, charts, and illustrations that may not translate well into Braille or large-print versions of the test. For 2006, the Writing/Speculate prompt from Language Arts Literacy, four items from Mathematics, and five items from Science were removed from the Braille form.

TABLE 2.11
Operational Test Specifications

Content Areas	Cluster	Number of Items		
		MC	OE	Total
<i>Language Arts Literacy</i>	<i>Reading</i>	20	6	26
	<i>Writing</i>	20	4	24
	<i>Writing/Speculate</i>		1	1
	<i>Writing/Persuade</i>		1	1
<i>Mathematics</i>	<i>Number and Numerical Operations</i>	30	6	36
	<i>Geometry and Measurement</i>	6	2	8
	<i>Patterns and Algebra</i>	9	1	10
	<i>Data Analysis, Probability, and Discrete Mathematics</i>	9	1	10
		6	2	8
<i>Science</i>	<i>Life</i>	45	3	48
	<i>Physical</i>	19	1	20
	<i>Earth</i>	13	1	14
		13	1	14

CHAPTER 3: TEST ADMINISTRATION

3.1 Participation

In 1988, the New Jersey State Legislature passed a law (18A:7C-6.2) requiring that a test be given to all eighth-grade students in public schools in New Jersey to assess their progress toward mastering the skills they will need to graduate from high school. All eighth-grade public school students must take the GEPA. This includes:

- General education students
- Limited-English Proficient (LEP) students
- Special Education (SE) students
- Students with Disabilities (Section 504)
- Retained eighth graders

In accordance with the Individuals with Disabilities Education Act (IDEA), students who are receiving special education services must participate in each subject area of the age-appropriate statewide assessment with the following exception:

Students with disabilities shall participate in the Alternate Proficiency Assessment in each content area where the nature of the student's disability is so severe that the student is not receiving instruction in any of the knowledge and skills measured by the general statewide assessment and the student cannot complete any of the types of questions on the assessment in the content area(s) even with accommodation and modifications. (New Jersey Administrative Code Chapter 6A:14-4.11[a]2)

The Alternate Proficiency Assessment (APA) is a portfolio-style assessment designed to measure progress toward achieving New Jersey's state educational standards for those students with severe disabilities who are unable to participate in the New Jersey Assessment of Skills and Knowledge (NJASK), the Grade Eight Proficiency Assessment (GEPA), or the High School Proficiency Assessment (HSPA).

3.2 Test Security Procedures

The test booklet and its contents are secure materials. They are not to be read or copied, either wholly or in part, for any purpose without express written permission from the New Jersey Department of Education. It is the responsibility of the school districts to guarantee the security of the test materials. Security breaches may have financial consequences for the district, professional consequences for staff, and disciplinary consequences for students.

The items and passages contained in the test booklet must remain confidential because some test items reappear in future versions of the tests. The answer folders (approximately 56 pages) contain grids for marking the answers to multiple-choice questions. Also, the answer folders are used by students for writing responses to the open-ended questions and the writing essay prompts. The security of test items and passages is required to maintain the stability of the test item pool over time from a technical perspective and to enable comparisons to be made from one year to the next. Examiners, proctors, and other school personnel are prohibited from discussing or disclosing any test items before, during, or after the test administration.

The following are secure materials for the GEPA administration:

- Test booklets
- Used answer folders
- All other answer folders until after testing
- Mathematics Reference Sheets until after testing

Pearson Educational Measurement (PEM) assigns a unique identification number to each secure test booklet and answer folder. The unique identification numbers are listed on security checklists. The unique identification number appears as a bar-code on test booklets. Following the test administration, PEM compares bar-code scan files of returned test booklets with distribution files to determine if all secure materials have been returned from each school and district. PEM contacts any district with missing secure test booklets or answer folders. For the 2006 administration, PEM scanned more than 120,000 secure test booklets.

The NJDOE Office of Evaluation and Assessment outlined the following security procedures in the *GEPA Test Manual*. District test coordinators were trained in these procedures during regional meetings held by the Office of Evaluation and Assessment in January and February 2006.

1. The chief school administrator or designee must sign for the initial shipment of test materials after presenting the Authorization to Receive Secure Test Materials form to the agent of the delivery service when the materials are delivered.
2. When not being used during testing, test materials must be stored in a secure, locked place that is accessible only to individuals whose access has been authorized by the school test coordinator. During testing, secure materials must not be removed from the testing room for review or photocopying. **Security of test materials must be maintained at all times.**
3. Each test booklet and answer folder has a **unique identification number**. Students must use the same test booklet and the same answer folder for each day of testing. On the first day of testing, students should print their name on the front cover of the test booklet assigned to them, and record the number and form letter of that test booklet on their answer folder.

4. Teachers are NOT to be given their own test booklet. The shrink-wrapped packaging on the test booklets may be opened for distribution just prior to testing.
5. Each day's section of the test booklet is sealed on all open sides. There are separate seals for the Science section, the Mathematics section, and Day 1 and Day 2 of the Language Arts Literacy section of the test. These seals must not be broken until the student breaks them the day that test section is administered.
6. District and school test coordinators must use the District and School Security Checklists to maintain an accurate record of the chain of distribution and collection of all test booklets.
7. Answer folders must not be duplicated or handscored.
8. An answer folder must be gridded for every enrolled Grade 8 student regardless of APA status.
9. An Irregularity Report form is used to report irregularities involving test booklets, answer folders, or anything that could impact test takers.
10. The principal and the chief school administrator or his/her designee must review and sign the completed Header sheets before they are submitted for scoring. The signatures affirm that the number of answer folders returned is correct and that all GEPA test administration procedures outlined in the manuals have been followed.
11. The Office of Evaluation and Assessment, in cooperation with county offices, monitors all aspects of testing and the implementation of security procedures at selected sites. Announcements of security visits are not made in advance.

The district test coordinators' training and the *Test Manual* include responsibility descriptions for the district test coordinator, school test coordinator, and examiner.

A security plan sample in the *Test Manual* delineated tasks and responsibilities for the following: turnkey training, storage of secure materials, delivery problems, missing test booklet, chain of command, sick child, disruptive student, fire drill/bomb scare, and inclement weather.

The Office of Evaluation and Assessment staff members monitor the test administration with specific procedures such as:

- Prior to actual testing, observe initial instructions to the students from the examiners and proctors.
- Observe all testing sites, including rooms where special accommodations are provided.

Breach test forms and examiner's manuals were prepared in the event of a security breach. In schools with the security breaches, appropriate staff members completed each student's name, date of birth, and answer folder number so that the alternate scoring occurred properly for the students. Specialized scoring and reporting included developing alternate test score keys, conversion tables, and reports.

3.3 Test Administration Procedures

The district test coordinators, school test coordinators, and examiners are responsible for the proper administration of the test. The district test coordinator is responsible for ensuring that examiners are selected and trained. All examiners must be certified teachers currently employed by the school. The district and school test coordinators, and examiners must read the *Test Manual* and *Examiner Manual* carefully to get an overview of all activities.

Student Rosters with appropriate Special Codes must be prepared to include each and every eighth-grade student in the district. The information from the rosters is used to code the “School Use Only” section of the student information grid on page one of the answer folder; to verify correct gridding by the student; or to verify the pre-ID label, if applicable.

The Student Rosters must:

- List each eighth-grade student’s name, date of birth, gender, and ethnicity
- Identify students with SE classifications, IEP exemptions/accommodations, or Section 504 status
- Identify students who are designated Title 1, economically disadvantaged, Limited English Proficient, and/or migrant status
- Designate coding for student’s time in district/time in school less than one year

Information from the Student Rosters is used to:

- ensure students are testing in the correct room
- code the “School Use Only” section of the student information grid on the answer folder
- verify correct gridding by students, and to
- verify that correct data appears on the pre-ID label for districts using labels

Test booklets and answer folders are distributed to examiners only on the morning of each day of the test administration. Specific instructions for the test administration are contained in the *Examiner Manual*. The examiners’ familiarity with the materials and the prescribed procedures is essential to the successful administration of the test. During the examiners’ training, district and school test coordinators emphasize that students can be given no assistance or coaching beyond what is specified in the manual.

When more than 25 students are tested in one room, the examiner uses the assistance of proctors. The school test coordinator briefs the proctors on the test materials and procedures, and specifies their responsibilities before, during, and after test administration. Proctors help in distributing and collecting non-secure materials, in observing students from different points in the room during test administration, and in answering student questions when there is a problem related to the test directions.

Total testing time (including time for distributing and collecting materials, reading directions, and taking breaks) is approximately nine hours over four successive days. The GEPA test administration must be scheduled in the morning. The Science, Mathematics, and Language Arts Literacy content-area tests were administered on the specified dates during the regular and make-up testing weeks.

3.4 Test Accommodations

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.

General education students receive no special testing accommodations other than the standard room setup and materials distribution described in the examiner's section of the *Test Manual*.

Limited English Proficient (LEP) students are tested with one or more of these accommodations:

- Additional time up to 150% of the administration times indicated
- Translation of directions only to the student's native language. Translations of passages, items, prompts, and tasks are NOT permitted
- Use of a bilingual dictionary, preferably one normally used by the student as part of the instructional program

Special education (SE) students must take the GEPA unless their Individualized Education Program (IEP) specifically states that they must be taking the Alternate Proficiency Assessment (APA) and not the GEPA.

Students with disabilities eligible for special education and related services and those students eligible under Section 504 of the Rehabilitation Act of 1973 may have accommodations and/or modifications during administration of the statewide assessment. Any accommodations or modifications of test administration procedures must be specified in the student's IEP or 504 accommodation plan. Accommodations or modifications must be consistent with the instruction and assessment procedures used in the student's classroom.

Section 504 students eligible for modifications may not be classified as special education but do have a permanent or temporary impairment in a major life function (for example: performing manual tasks, walking, seeing, hearing, speaking, etc.). A Section 504 plan may be permanent or temporary. A student with a temporary/emergency Section 504 Plan must be identified as 504 and have the appropriate accommodations identified. These students must be tested using modified testing procedures that are specified in the student's 504 accommodation plan and that are approved by the Department of Education. The temporary/emergency plan allows the student to receive all necessary testing accommodations and modifications, such as additional time and the use of a computer or scribe.

Accommodations and modifications of test administration procedures are listed in Appendix C of this report. Also, the accommodations and modifications are included in the *Test Manual*, the *Examiner Manual*, and at <http://www.state.nj.us/njded/specialed/accom900.htm>

If a student requires an accommodation or modification that is not listed, district staff are instructed to contact the Office of Evaluation and Assessment, GEPA Coordinator. Accommodations or modifications are classified as follows:

- A = Setting Accommodations
- B = Scheduling Accommodations
- C = Test Materials/Modifications
- D = Test Procedures Modifications

Visually impaired students may take either a Braille or large-print version of the test. Specific instructions for administering the Braille and large-print versions of the test are provided in the supplementary instructions for examiners administering these forms.

Students using the Braille test booklets:

- are instructed to bring a Braille ruler and a talking calculator to the test session.
- are instructed to skip some items identified in the Braille instructions. The spaces for these items must be left blank on the student answer folder.
- have answer folders transcribed from Braille version by the examiner.
- dictate their answers to the examiner or use a device that produces Braille. For dictations and responses recorded in Braille:
 - Students must indicate all punctuation and must spell all key words.
 - Examiners must transcribe the Brailled responses into the regular answer folder.

Students using the large-print test booklets:

- mark their answers in the large-print answer folders.
- may be instructed to skip some questions. The spaces for these questions must be left blank in the student's large-print answer folder.
- who dictate responses on open-ended items and writing tasks indicate all punctuation and spell key words.

3.5 Results for Special Education Students and Section 504 Students Tested with Accommodations or Modifications

The following tables show the proficiency level results for special education students and Section 504 students tested with accommodations and modifications. Also, the first row of each table includes the number of students and performance results for Special Education students as shown in Table 1.2 of this *Technical Report* and the state level Performance by Demographic Groups Report from Cycle II reporting.

Not every special education student or Section 504 student is tested with an accommodation or modification. Accommodations and modifications may be used separately or in combination. The table below shows the number of special education students with performance results and the number of Section 504 students with performance results tested with each of the accommodations and modifications.

The tables on the following pages show the numbers of students and proficiency results by special education disability category. Instructions to the examiners note that "...one and only one disability category for each special education student..." should be designated. The N category is used to indicate multiple grids. Also, the N category is a default code used when a school fails to provide the specific disability-category information listed for an APA student.

TABLE 3.1
Setting Accommodations and Modifications for Special Education Students and Section 504 Students

	NUMBER OF STUDENTS ENROLLED	LANGUAGE ARTS LITERACY						MATHEMATICS						SCIENCE								
		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)				
		NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%			
Special Education	18,327	17,076	11,491	67.3	5,493	32.2	92	0.5	17,279	12,955	75.0	3,791	21.9	533	3.1	17,333	8,702	50.2	7,787	44.9	844	4.9
Special Education Accommodations		14,723	10,036	68.2	4,611	31.3	76	0.5	14,951	11,325	75.7	3,203	21.4	423	2.8	14,997	7,592	50.6	6,748	45.0	657	4.4
Setting Accommodations (A)		14,631	9,944	68.0	4,609	31.5	78	0.5	14,876	11,231	75.5	3,220	21.6	425	2.9	14,920	7,545	50.6	6,703	44.9	672	4.5
Scheduling Accommodations (B)		792	494	62.4	292	36.9	6	0.8	741	519	70.0	195	26.3	27	3.6	743	311	41.9	379	51.0	53	7.1
Test Materials Modifications (C)		13,381	9,241	69.1	4,071	30.4	69	0.5	13,591	10,345	76.1	2,879	21.2	367	2.7	13,635	6,987	51.2	6,075	44.6	573	4.2
Test Procedures Modifications (D)																						
Section 504 Accommodations		1,362	333	24.4	951	69.8	78	5.7	1,363	433	31.8	686	50.3	244	17.9	1,367	256	18.7	865	63.3	246	18.0
Setting Accommodation (A)		1,479	364	24.6	1,032	69.8	83	5.6	1,484	471	31.7	741	49.9	272	18.3	1,489	274	18.4	938	63.0	277	18.6
Scheduling Accommodations (B)		92	17	18.5	66	71.7	9	9.8	78	24	30.8	35	44.9	19	24.4	78	9	11.5	43	55.1	26	33.3
Test Materials Modifications (C)		994	252	25.4	685	68.9	57	5.7	998	322	32.3	501	50.2	175	17.5	1,003	206	20.5	634	63.2	163	16.3
Test Procedures Modifications (D)																						

TABLE 3.2
Setting Accommodations for Special Education Students and Section 504 Students

	NUMBER OF STUDENTS ENROLLED	LANGUAGE ARTS LITERACY						MATHEMATICS						SCIENCE								
		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)				
		NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%			
Special Education	18,327	17,076	11,491	67.3	5,493	32.2	92	0.5	17,279	12,955	75.0	3,791	21.9	533	3.1	17,333	8,702	50.2	7,787	44.9	844	4.9
Setting Accommodations		14,723	10,036	68.2	4,611	31.3	76	0.5	14,951	11,325	75.7	3,203	21.4	423	2.8	14,997	7,592	50.6	6,748	45.0	657	4.4
A.Auditorily Impaired		121	75	62.0	44	36.4	2	1.7	125	81	64.8	36	28.8	8	6.4	125	65	52.0	54	43.2	6	4.8
B.Other Health Impaired		1,672	935	55.9	721	43.1	16	1.0	1,700	1,084	63.8	528	31.1	88	5.2	1,709	584	34.2	979	57.3	146	8.5
C.Communication Impaired		807	621	77.0	182	22.6	4	0.5	814	665	81.7	141	17.3	8	1.0	816	528	64.7	280	34.3	8	1.0
D.Emotionally Disturbed		1,050	796	75.8	248	23.6	6	0.6	1,089	881	80.9	173	15.9	35	3.2	1,094	595	54.4	454	41.5	45	4.1
E.Cognitively Impaired		357	340	95.2	17	4.8	0	0.0	357	340	95.2	16	4.5	1	0.3	363	324	89.3	38	10.5	1	0.3
F.Multiply Disabled		1,459	1,230	84.3	223	15.3	6	0.4	1,479	1,286	87.0	163	11.0	30	2.0	1,482	975	65.8	453	30.6	54	3.6
G.Traumatic Brain Injury		83	65	78.3	17	20.5	1	1.2	80	67	83.8	13	16.3	0	0.0	80	54	67.5	25	31.3	1	1.3
H.Orthopedically Impaired		54	25	46.3	28	51.9	1	1.9	54	34	63.0	17	31.5	3	5.6	53	18	34.0	33	62.3	2	3.8
I.Specific Learning Disability		8,865	5,778	65.2	3,052	34.4	35	0.4	9,004	6,709	74.5	2,066	22.9	229	2.5	9,026	4,327	47.9	4,330	48.0	369	4.1
J.Social Maladjustment		16	12	75.0	4	25.0	0	0.0	16	14	87.5	2	12.5	0	0.0	17	7	41.2	10	58.8	0	0.0
K.Visually Impaired		15	7	46.7	8	53.3	0	0.0	15	5	33.3	1	6.7	0	0.0	5	1	20.0	3	60.0	1	20.0
L.Speech-Language Services Only		52	22	42.3	30	57.7	0	0.0	52	33	63.5	12	23.1	7	13.5	52	18	34.6	31	59.6	3	5.8
M.Autistic		143	104	72.7	35	24.5	4	2.8	147	100	68.0	33	22.4	14	9.5	146	76	52.1	49	33.6	21	14.4
N.Multiple Grids and Default		29	26	89.7	2	6.9	1	3.4	29	27	93.1	2	6.9	0	0.0	29	20	69.0	9	31.0	0	0.0
Section 504		1,362	333	24.4	951	69.8	78	5.7	1,363	433	31.8	686	50.3	244	17.9	1,367	256	18.7	865	63.3	246	18.0

TABLE 3.3
Scheduling Accommodations for Special Education Students and Section 504 Students

	NUMBER OF STUDENTS ENROLLED	LANGUAGE ARTS LITERACY						MATHEMATICS						SCIENCE							
		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)			
		NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%		
Special Education	18,327	11,491	67.3	5,493	32.2	92	0.5	17,279	12,955	75.0	3,791	21.9	533	3.1	17,333	8,702	50.2	7,787	44.9	844	4.9
Scheduling Accommodations	14,631	9,944	68.0	4,609	31.5	78	0.5	14,876	11,231	75.5	3,220	21.6	425	2.9	14,920	7,545	50.6	6,703	44.9	672	4.5
A. Auditorily Impaired	120	75	62.5	44	36.7	1	0.8	124	81	65.3	36	29.0	7	5.6	124	65	52.4	54	43.5	5	4.0
B. Other Health Impaired	1,667	938	56.3	713	42.8	16	1.0	1,694	1,077	63.6	527	31.1	90	5.3	1,703	582	34.2	969	56.9	152	8.9
C. Communication Impaired	801	613	76.5	184	23.0	4	0.5	808	658	81.4	143	17.7	7	0.9	810	521	64.3	281	34.7	8	1.0
D. Emotionally Disturbed	1,049	794	75.7	249	23.7	6	0.6	1,094	884	80.8	176	16.1	34	3.1	1,099	601	54.7	451	41.0	47	4.3
E. Cognitively Impaired	346	329	95.1	17	4.9	0	0.0	346	329	95.1	16	4.6	1	0.3	351	312	88.9	38	10.8	1	0.3
F. Multiply Disabled	1,464	1,235	84.4	223	15.2	6	0.4	1,492	1,301	87.2	160	10.7	31	2.1	1,493	990	66.3	449	30.1	54	3.6
G. Traumatic Brain Injury	84	66	78.6	17	20.2	1	1.2	85	68	80.0	13	16.0	0	0.0	81	55	67.9	25	30.9	1	1.2
H. Orthopedically Impaired	45	23	51.1	21	46.7	1	2.2	45	30	66.7	14	31.1	1	2.2	44	17	38.6	25	56.8	2	4.5
I. Specific Learning Disability	8,809	5,709	64.8	3,063	34.8	37	0.4	8,954	6,338	71.1	2,084	23.3	232	2.6	8,976	4,288	47.8	4,314	48.1	374	4.2
J. Social Maladjustment	16	12	75.0	4	25.0	0	0.0	16	14	87.5	2	12.5	0	0.0	17	7	41.2	10	58.8	0	0.0
K. Visually Impaired	16	8	50.0	8	50.0	0	0.0	5	4	80.0	1	20.0	0	0.0	5	1	20.0	3	60.0	1	20.0
L. Speech-Language Services Only	41	15	36.6	26	63.4	4	9.8	41	23	56.1	11	26.8	7	17.1	41	13	31.7	25	61.0	3	7.3
M. Autistic	143	101	70.6	38	26.6	4	2.8	146	97	66.4	34	23.3	15	10.3	146	73	50.0	50	34.2	23	15.8
N. Multiple Grids and Default	30	26	86.7	2	6.7	2	6.7	30	27	90.0	3	10.0	0	0.0	30	20	66.7	9	30.0	1	3.3
Section 504	1,479	364	24.6	1,032	69.8	83	5.6	1,484	471	31.7	741	49.9	272	18.3	1,489	274	18.4	938	63.0	277	18.6

TABLE 3.4
Test Materials Modifications for Special Education Students and Section 504 Students

	NUMBER OF STUDENTS ENROLLED	LANGUAGE ARTS LITERACY						MATHEMATICS						SCIENCE							
		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)		PARTIALLY PROFICIENT (100 – 199)		PROFICIENT (200 – 249)		ADVANCED PROFICIENT (250 – 300)			
		NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%		
Special Education	18,327	11,491	67.3	5,493	32.2	92	0.5	17,279	12,955	75.0	3,791	21.9	533	3.1	17,333	8,702	50.2	7,787	44.9	844	4.9
Test Materials Modifications	792	494	62.4	292	36.9	6	0.8	741	519	70.0	195	26.3	27	3.6	743	311	41.9	379	51.0	53	7.1
A. Auditorily Impaired	5	2	40.0	3	60.0	0	0.0	5	2	40.0	3	60.0	0	0.0	5	2	40.0	3	60.0	0	0.0
B. Other Health Impaired	107	56	52.3	50	46.7	1	0.9	106	68	64.2	30	28.3	8	7.5	107	34	31.8	59	55.1	14	13.1
C. Communication Impaired	40	31	77.5	8	20.0	1	2.5	40	30	75.0	10	25.0	0	0.0	39	23	59.0	16	41.0	0	0.0
D. Emotionally Disturbed	40	27	67.5	13	32.5	0	0.0	41	31	75.6	7	17.1	3	7.3	40	21	52.5	16	40.0	3	7.5
E. Cognitively Impaired	13	11	84.6	2	15.4	0	0.0	10	9	90.0	1	10.0	0	0.0	10	8	80.0	2	20.0	0	0.0
F. Multiply Disabled	91	71	78.0	20	22.0	0	0.0	71	58	81.7	12	16.9	1	1.4	70	36	51.4	30	42.9	4	5.7
G. Traumatic Brain Injury	6	5	83.3	1	16.7	0	0.0	2	2	100.0	0	0.0	0	0.0	2	1	50.0	1	50.0	0	0.0
H. Orthopedically Impaired	7	1	14.3	6	85.7	0	0.0	6	0	0.0	5	83.3	1	16.7	6	0	0.0	5	83.3	1	16.7
I. Specific Learning Disability	450	273	60.7	173	38.4	4	0.9	440	309	70.2	119	27.0	12	2.7	443	179	40.4	238	53.7	26	5.9
J. Social Maladjustment	0	0	0.0	0	0.0	0	0.0	0	0	0.0	0	0.0	0	0.0	1	0	0.0	1	100.0	0	0.0
K. Visually Impaired	14	6	42.9	8	57.1	0	0.0	2	1	50.0	1	50.0	0	0.0	2	0	0.0	1	50.0	1	50.0
L. Speech-Language Services Only	3	1	33.3	2	66.7	0	0.0	3	1	33.3	2	66.7	0	0.0	3	0	0.0	2	66.7	1	33.3
M. Autistic	16	10	62.5	6	37.5	0	0.0	15	8	53.3	5	33.3	2	13.3	15	7	46.7	5	33.3	3	20.0
N. Multiple Grids and Default	0	0	0.0	0	0.0	0	0.0	0	0	0.0	0	0.0	0	0.0	0	0	0.0	0	0.0	0	0.0
Section 504	92	17	18.5	66	71.7	9	9.8	78	24	30.8	35	44.9	19	24.4	78	9	11.5	43	55.1	26	33.3

TABLE 3.5
Test Procedures Modifications for Special Education Students and Section 504 Students

	NUMBER OF STUDENTS ENROLLED	LANGUAGE ARTS LITERACY						MATHEMATICS						SCIENCE								
		PARTIALLY PROFICIENT (100 - 199)		PROFICIENT (200 - 249)		ADVANCED PROFICIENT (250 - 300)		PARTIALLY PROFICIENT (100 - 199)		PROFICIENT (200 - 249)		ADVANCED PROFICIENT (250 - 300)		PARTIALLY PROFICIENT (100 - 199)		PROFICIENT (200 - 249)		ADVANCED PROFICIENT (250 - 300)				
		NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%	NO.	%			
Special Education	18,327	17,076	11,491	67.3	5,493	32.2	92	0.5	17,279	12,955	75.0	3,791	21.9	533	3.1	17,333	8,702	50.2	7,787	44.9	844	4.9
Test Procedures Modifications	13,381	9,241	69.1	4,071	30.4	69	0.5	13,591	10,345	76.1	2,879	21.2	367	2.7	13,635	6,987	51.2	6,075	44.6	573	4.2	
A. Auditorily Impaired	114	69	60.5	42	36.8	3	2.6	118	75	63.6	34	28.8	9	7.6	118	61	51.7	49	41.5	8	6.8	
B. Other Health Impaired	1,473	844	57.3	615	41.8	14	1.0	1,499	965	64.4	458	30.6	76	5.1	1,507	527	35.0	856	56.8	124	8.2	
C. Communication Impaired	747	586	78.4	157	21.0	4	0.5	754	621	82.4	127	16.8	6	0.8	756	499	66.0	249	32.9	8	1.1	
D. Emotionally Disturbed	925	705	76.2	215	23.2	5	0.5	963	785	81.5	148	15.4	30	3.1	968	530	54.8	397	41.0	41	4.2	
E. Cognitively Impaired	346	332	96.0	14	4.0	0	0.0	345	330	95.7	14	4.1	1	0.3	350	314	89.7	36	10.3	0	0.0	
F. Multiply Disabled	1,377	1,173	85.2	199	14.5	5	0.4	1,398	1,225	87.6	147	10.5	26	1.9	1,405	935	66.5	423	30.1	47	3.3	
G. Traumatic Brain Injury	80	64	80.0	15	18.8	1	1.3	77	65	84.4	12	15.6	0	0.0	77	54	70.1	22	28.6	1	1.3	
H. Orthopedically Impaired	53	25	47.2	27	50.9	1	1.9	53	35	66.0	15	28.3	3	5.7	52	19	36.5	31	59.6	2	3.8	
I. Specific Learning Disability	8,053	5,307	65.9	2,715	33.7	31	0.4	8,177	6,105	74.7	1,876	22.9	196	2.4	8,195	3,953	48.2	3,927	47.9	315	3.8	
J. Social Maladjustment	16	12	75.0	4	25.0	0	0.0	16	14	87.5	2	12.5	0	0.0	16	7	43.8	9	56.3	0	0.0	
K. Visually Impaired	13	7	53.8	6	46.2	0	0.0	4	3	75.0	1	25.0	0	0.0	4	1	25.0	2	50.0	1	25.0	
L. Speech-Language Services Only	49	22	44.9	27	55.1	0	0.0	49	31	63.3	11	22.4	7	14.3	49	15	30.6	31	63.3	3	6.1	
M. Autistic	132	93	70.5	35	26.5	4	3.0	135	89	65.9	33	24.4	13	9.6	135	71	52.6	42	31.1	22	16.3	
N. Multiple Grids and Default	3	2	66.7	0	0.0	1	33.3	3	2	66.7	1	33.3	0	0.0	3	1	33.3	1	33.3	1	33.3	
Section 504	994	252	25.4	685	68.9	57	5.7	998	322	32.3	501	50.2	175	17.5	1,003	206	20.5	634	63.2	163	16.3	

CHAPTER 4: SCORING

4.1 Multiple-choice Items

Each multiple-choice item contributes one point to the total raw score for each content-area test. Responses for multiple-choice items are machine scored. The score points of multiple-choice items received for a content area are the total number of multiple-choice items answered correctly. For the Mathematics and Science content areas and the Language Arts reading component, the total score points of multiple-choice items are combined with the total number of points from the open-ended items for a student's score. For Language Arts Literacy, the reading component score points are added to score points received from the open-ended scoring of the two writing tasks which compose the writing component.

4.2 Open-ended Items

During April and May of 2006, Measurement Incorporated (MI) under subcontract to Pearson Educational Measurement (PEM) scored the student writing responses, and the reading, mathematics, and science open-ended items. MI has a staff of highly-trained scorers who must have at least a bachelor's degree and who must undergo rigorous and ongoing training and monitoring during the scoring process. Each open-ended item and each writing prompt was read independently by two scorers. If the two scorers disagreed by more than one point, a third scorer evaluated the response. Appendix A presents information about how the three scores are resolved for each of the content areas.

Table 4.1 shows the number of writing responses and open-ended items scored for the operational test.

TABLE 4.1

Number of Writing Prompts and Open-ended Items Scored

Content Area	Number of Writing Prompts and Open-ended Items Scored
<i>Language Arts Literacy</i>	1,320,922
<i>Reading</i>	879,800
<i>Writing</i>	441,122
<i>Speculate</i>	220,307
<i>Persuade</i>	220,815
<i>Mathematics</i>	1,318,241
<i>Science</i>	665,018
TOTAL	3,304,181

Scorer Selection

MI's senior project managers work closely with Content Coordinators in the Office of Evaluation and Assessment. Current procedures for scoring the GEPA open-ended and writing responses are consistent with those used since the inception of a performance-based writing component in the New Jersey statewide assessment. Scoring of the open-ended and writing responses is monitored by trained, experienced personnel who have met the same rigorous standards established with the initial holistic scoring study conducted in 1986.

For selecting team leaders, MI's management staff and scoring directors reviewed the files of all returning staff who have previously scored the GEPA. The MI staff looked for people who were experienced team leaders with a record of good performance on previous projects and also considered scorers who have been recommended for promotion to the team leader position.

Many of the MI scorers have repeatedly scored the GEPA for previous test administrations. MI's procedures for selecting new scorers are very thorough. After advertising in local newspapers, with the job service, and elsewhere, and receiving applications, staff in MI's human resources department review applications and schedule interviews for qualified applicants. Qualified applicants are those with a four-year college degree in English, language arts, education, mathematics, science, or a related field. Each qualified applicant must pass an interview by experienced MI staff, write an acceptable essay, and receive good recommendations from references. All the information about each applicant is reviewed before offering employment.

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

Rangefinding

Rangefinding is one of the most important elements of the scoring process. Rangefinding meetings provide an opportunity for finalizing scoring rubrics (in content areas with specific item rubrics) and making scoring decisions and interpretations regarding scoring issues before team leader and scorers' training begins. (See Appendix A for rubrics.) It is important that as many of the item-specific problems as possible be resolved prior to scorers' training so that scoring decisions can be made during scoring.

After consulting with PEM to determine when the first “live” student responses would be available, MI scheduled a rangefinding meeting in Durham, other MI sites (operational test), and New Jersey (field test) to establish “true” scores for a representative sample of open-ended items. At this meeting, Office of Evaluation and Assessment staff members, content committee members, and the MI project leaders read and scored 60-225 responses, which exemplified various points of the rubric and score scale. The number of responses varied according to the content area and score scale. The responses were selected from a broad range of New Jersey school districts in order to ensure that the sample was representative of overall student performance. Rangefinding took from two to six days per content area, depending on the number of items tested.

Development of Scoring Guides

After the rangefinding responses were discussed and received a final score, MI used the selected responses to develop scoring guides, training sets (practice papers), and/or qualifying sets for each content area. Scoring guides consisted of three or more examples of each score point in score point order. In some content areas, the papers were annotated. Training and qualifying sets were clearly anchored papers in random score point order. Sufficient copies were made so that all scoring directors, team leaders, and scorers had their own copy during training and scoring.

Team Leader Training and Qualifying

After the anchor papers, training, and/or qualifying papers were identified and finalized, team leader training began. The scoring director (for each content area or writing type) conducted training for the team leaders. Procedures were similar to those for training scorers (see below) but were more comprehensive, dealing with resolution of discrepant scores, identification of nonscorable responses, unusual prompt treatment, alert situation responses (e.g., child-in-danger), and other duties performed only by team leaders. The team leaders carefully prepared notes on the training papers in preparation for discussion with the scorers, and the scoring director counseled team leaders on training techniques and application of the rubric.

Team leaders assisted in training scorers in team discussions of training sets, and were responsible for distributing, collecting, and accounting for training packets and sample papers during each scoring session. During scoring, team leaders responded to questions, spot-checked reader packets, and counseled scorers having difficulty with the criteria.

Team leaders also administered the quality control (validity) set, monitored the scoring patterns of each reader throughout the project, conducted retraining as necessary, performed some resolution readings, and maintained a professional working environment. The validity sets were generally selected by the team leaders and scoring director for each content area prior to reader training.

Team leader training lasted from two to four days. Team leaders generally worked 7.75 hours per day, excluding breaks. They set up the room prior to reader arrival each day and met with scoring directors after scoring each day.

Scorer Training and Qualifying

All scorers were trained using the scoring guides and rubrics, training papers, and/or qualifying papers selected during the rangefinding meetings. Scorers were assigned to a scoring group consisting of one team leader and 10-12 scorers. Each scorer was assigned an individual number for easy identification of their scoring work throughout the scoring session.

After the contracts and nondisclosure forms were signed and the introductory remarks given, training began. Scorer training followed the same format as team leader training except that scorers were not required to annotate each paper in the training sets, although they were encouraged to take notes. The scoring director presented the writing or open-ended item task and introduced the guide, then discussed, room-wide, each score point. This presentation was followed by practice scoring on the training sets. Each scorer was given a clean copy of the scoring guide and training sets, as well as a monitor sheet on which to record training set scores. Because it is easy in a large group to overlook a shy scorer who may be having difficulty, scorers did break into teams to score and discuss the papers in the training sets. This arrangement provided scorers an opportunity to discuss any possible points of confusion or problems in understanding the criteria.

Team leaders collected the monitor sheets after the scoring of each training set and recorded results on a customized log, which was examined by the scoring director to determine which papers were giving scorers difficulty. The scoring director also “floated” from team to team, listening to the team leaders’ explanations and adding additional information when necessary. If a particular paper or type of paper seemed to cause difficulty across teams, the problem was discussed room-wide to ensure that everyone heard the same explanation.

Like team leaders, scorers must demonstrate their ability to score accurately by attaining the agreement percentage established by the New Jersey Department of Education before they may score packets of “live” papers. Any scorer unable to meet these standards was dismissed. All scorers understand this stipulation when they are hired.

Training was carefully orchestrated so that scorers understood how to apply the rubric in scoring the papers, learned how to reference the scoring guide, developed the flexibility needed to deal with a variety of responses, and retained the consistency needed to score all papers accurately.

Scorers were trained to recognize and flag nonscorable responses (fragment, off-topic, not English, no response) and “alert” papers (e.g., suspicion of child abuse) so that these papers could be handled in the correct manner. Alert papers were scored, but then forwarded to the scoring director for review. If the scoring director agreed that the student’s own words specifically stated a situation that qualified as an alert or reflected a potential risk situation for a child, the paper was copied and sent to the Office of Evaluation and Assessment for follow-up with school district personnel. Alert papers are flagged if they reflect potential abuse, emotional or psychological difficulty, dangerous thoughts, or possible plagiarism.

In addition to completing all of the initial training and qualifying, a significant amount of time was allotted for demonstrations of paper flow, explanations of “alerts” and “flagging,” and instructions about other procedures which were necessary for the conduct of a smooth project. Scorer training lasted from two to five days. Scorers generally worked 7.0 hours per day, excluding breaks.

Scoring Procedures and Paper Flow

Each student response was scored by two independent scorers using the scoring scale developed and approved for those items. If the two assigned scores differed by more than one point, the paper was returned for a third “resolution” reading by team leaders or scoring directors. Information about how the three scores were resolved appears in Appendix A.

Before opening a packet, scorers began by writing their assigned reader numbers, as well as the date, on the front of their packet envelope. The stapled packet of papers and the appropriate monitor sheet (first or second reading) was then removed from the envelope. Scorers checked the packet number on the header sheet against the number on the monitor sheet for agreement, and then recorded their scorer identification numbers in the designated space on the scannable monitor sheet. The scorer decided on the score, and the assigned scores are recorded in the appropriate spaces provided on the monitor. As scorers progressed through a packet, they checked each paper’s student ID number against the number printed on the monitor sheet. If there was a discrepancy, the packet was flagged for the scoring director to check.

As a scorer completed a packet of papers, he or she returned it to the envelope and gave it to the team leader, along with the monitor sheet. The clerical aide picked up completed packets and monitor sheets, and redistributed the packets for second readings.

The packet proceeded to the second reading stage while the first reading scores were being scanned. The procedure for the second reading was the same as that for the first reading, except that the second scorer used the second scoring monitor sheet in the envelope. At no time does the second scorer have access to the scores given by the first scorer. As with the first scoring monitors, the second monitors were scanned and the scores merged into the database.

After the second scores were entered, they were matched with the first scores already in the database. When scores differed by more than one point on any response, the response was classified as “discrepant,” a third scoring list by packet and response number was printed, and the response was returned for a third independent reading. After the clerical aide returned the packet to the scoring room, the scoring director located the papers needing a third reading and followed the normal scoring procedures. The third score was scanned in the same manner as the first two scores. The packet was returned to the warehouse and refiled.

Scorer Monitoring

Scorers were monitored in several ways. Team leaders answered scorers' questions, using the guide and training papers as examples. They also read behind their team members by reviewing packets after they were turned in, looking for papers that might merit discussion with the scorer. In addition, every day the scoring director and team leaders received the printout of the scorer statistics—including the scorers' perfect, adjacent, and resolution agreement with other scorers, and the scorers' score point distribution. In this way, the scoring director and team leader can look at any one scorer, team, or the room as a whole and rollover items can be compared to previous years.

Agreement Between Scorers for the Writing Tasks and Open-Ended Items

Table 4.2 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:

- writing/speculate task in response to a picture —
1-6 points, scorer ratings averaged
- writing/persuade task —
1-6 points, scorer ratings summed

Each writing task is rated by two independent scorers. Of the more than 220,000 task responses scored for the 2006 administration, 63.0% received exactly the same scores by the scorers and 35.2% received scores that were adjacent. Thus, approximately 98.2% of the task responses required only two scorers. The remaining 1.9% received scores on the writing tasks that differed by more than one point and, therefore, required a third scorer.

All content areas included open-ended items. For the Reading open-ended items, the rubric used by the scorers had score points that ranged from 0 to 4. Two Reading open-ended items are presented for each of two reading passages. For these four items, the resolution percent ranged from 1.0% to 1.8% with the percent at perfect agreement ranging from 62.1% to 68.5%.

Six open-ended items were presented for Mathematics. These six items had percents at perfect agreement ranging from 81.5% to 92.6%. The percent requiring resolution ranged from 0.5% to 2.0%.

Three open-ended items were included for Science. These items had a perfect agreement rate ranging from 77.7% to 80.7%. The percent requiring resolution ranged from 1.1% to 6.1%.

4.3 Quality Control Procedures in Data Preparation

Quality control procedures at Pearson Educational Measurement (PEM) begin with the use of the Capability Maturity Model (CMM), a software development management tool. Key process areas of CMM are requirements management, software project planning, software project tracking and oversight, software quality assurance, and software configuration management. PEM examples of CMM documents include a customer requirements allocation document, a project schedule, functional specifications, a software development project plan, unit test plans, and verification and validation plans. PEM is certified by an external auditor for CMM Level 4, the second highest level of certification.

TABLE 4.2
Consistency Between Raters Scoring
GEPA Writing Tasks and Open-Ended Items
March 2006

GEPA Writing Tasks and Open-Ended Items	Percent Raters In Exact Agreement	Percent Raters In Adjacent Agreement	Percent Resolution Needed
Language Arts Literacy			
Writing Total	63.0	35.2	1.9
Writing/Speculate	64.0	34.4	1.6
Writing/Persuade	61.9	36.0	2.1
Reading Total	65.8	32.9	1.3
Open-Ended Item 1	67.0	32.0	1.0
Open-Ended Item 2	68.5	30.3	1.1
Open-Ended Item 3	62.1	36.1	1.8
Open-Ended Item 4	65.7	33.0	1.4
Mathematics			
Mathematics Total	86.5	12.4	1.1
Open-Ended Item 1	82.7	16.0	1.2
Open-Ended Item 2	87.3	11.5	1.1
Open-Ended Item 3	81.5	16.5	2.0
Open-Ended Item 4	89.8	9.7	0.5
Open-Ended Item 5	85.2	13.9	0.9
Open-Ended Item 6	92.6	6.6	0.8
Science			
Science Total	78.7	18.4	2.9
Open-Ended Item 1	80.7	17.9	1.4
Open-Ended Item 2	77.7	21.2	1.1
Open-Ended Item 3	77.8	16.1	6.1

After software requirements have been identified, the PEM software development team prepares project schedules, project plans, functional specifications, and design documents. PEM begins by creating detailed test plans at both the unit and systems level. A unit test plan is a list of code-unit test cases that are executed and recorded by the software developer. The purpose of the code-unit test process is to ensure that software is developed, maintained, documented, and verified to meet the project requirements for coding and unit testing. As such, the process provides the mechanisms that are necessary to implement the software requirements and design as well as provides code-units quality assurance prior to system test.

After all modules (units) are tested within a system, the CMM process requires a system test. The system test ensures that all the units work together and that outputs from one module match up to the proper inputs for the next module in the system. It also uses expected results to ensure that all requirements have been met. It is important that the system test be performed by a group that is independent of the software development team. This process allows independent verification and interpretation of the requirements. Once the independent testing group has completed the test and given its approval, the system is moved into production mode. It is ready for processing the quality-checking answer documents and files submitted by a quality-checking team.

Scanning and Scoring

Before actual answer documents are machine-scanned, a comprehensive check of the scanning and scoring system is performed. The software development tester creates test decks of gridded answer documents with specific test criteria. The test decks are designed and gridded to cover all response ranges, ID ranges, blanks, and double grids as well as any other responses used by the GEPA. A file containing the scanned responses is then compared to the expected test results for each document to ensure the scanner is operating correctly. The test decks are processed through the programs for scanning and editing answer documents, and packetizing and printing scoring monitors.

The second check involves processing and quality-checking the first actual answer documents received. The NJDOE Office of Evaluation and Assessment and PEM asked 58 districts to return their answer documents early following the test administration so that all test forms could be processed and quality-checked. Also, these early return districts provided the actual student papers for determining score ranges for the writing tasks and open-ended items. Districts were selected to be representative for size and DFG. All information on approximately 60 answer documents was hand checked against the scanned file. In addition, periodically, throughout the processing of the documents, individual answer documents were checked by hand to ensure that scanning was continuing to perform correctly.

NJDOE Quality Control of Score Reporting

NJDOE Office of Evaluation and Assessment conducted the first round of quality control of multiple-choice items scoring on May 1-5, 2006, in New Jersey. PEM printed score sheets for each of approximately 1,000 students from more than 20 districts selected by the Office of Evaluation and Assessment for quality control.

Original answer folders for all students in the quality control sample were shipped to the meeting site. PEM maintained a copy of all answer folders in the quality control sample. PEM provided the following materials to the Office of Evaluation and Assessment for the quality control:

1. Scoring masks (punched index and transparency sheets) for all versions of the tests
2. Answer keys for the multiple-choice items
3. Double-grid documentation included a sample of edits for students who marked more than one answer for a multiple-choice item
4. Irregularity reports included all reports dealing with multiple answer folders for students and provided documentation about how these answer folders were merged
5. List of removed items from the Braille and large-print forms
6. List of names of all students taking a Braille or large-print form
7. County-district-school master files with district test coordinators' names and phone numbers
8. Frequency distributions for the student groups, including total, general, LEP, SE, IEP exempts by content area, void counts by reporting category, and Title 1 counts by reporting category

In the two weeks following the first round of quality control, Measurement Incorporated completed scoring the open-ended and essay responses. Assessment and Evaluation Services equated the test forms after which the NJDOE Office of Evaluation and Assessment and independent reviewers approved the equating procedures and raw score to scale score conversion tables. PEM staff loaded the conversion tables and produced Cycle I score reports for the quality control sample for review.

The second round of the Office of Evaluation and Assessment quality control on the Cycle I score reports took place on May 22-26, 2006, at PEM in Iowa City, Iowa. At this time the open-ended and essay scores were available.

The multiple-choice, open-ended, and essay item scores for each cluster and total for the three content areas were systematically checked on all Cycle I score reports. Individual Student Reports for all large-print, Braille, and breach students were produced and reviewed.

Calculations for the Total Scale Score Means and the Just Proficient Means (the mean score for all students across the state whose scale scores were 200 on a particular content area) were verified for each cluster in the content areas by the Office of Evaluation and Assessment staff. Summary statistics included on the School and District Summary Statistics reports were reviewed and approved.

CHAPTER 5: STANDARD SETTING

5.1 Overview of the Process

A proficiency level setting (standard setting) was conducted June 8-11, 1999, to describe and delineate the thresholds of performance that are indicative of Partially Proficient, Proficient, and Advanced Proficient performance for the GEPA Language Arts Literacy and Mathematics. A standard setting study for Science was conducted July 10-12, 2000. Results of these studies were used to formulate recommendations to the Commissioner of Education and the New Jersey State Board of Education for the adoption of the cut scores (i.e., proficiency levels).

The standard setting studies in 1999 and 2000 were conducted by staff from the New Jersey Department of Education, Office of Assessment; Assessment and Evaluation Services; and NCS Pearson. The document, *GEPA Standard Setting Report*, outlines the studies and presents the resulting documentation.

Participants in the standard setting study were chosen because of their qualifications as judges of student performance and content expertise. The judges represented the general population of New Jersey educators. Special care was taken to ensure adequate professional, gender, racial/ethnic, regional, and District Factor Group (DFG) representation on all panels.

A holistic classification method was used for the GEPA standard settings. The judges reviewed student papers sampled to represent the full range of student scores for the March 1999 GEPA administration of the Language Arts Literacy and Mathematics. The judges were asked to classify student work into three categories: Partially Proficient, Proficient, and Advanced Proficient. The judges had the opportunity to review, discuss, and modify their proficiency classifications. Using a logistic regression method, two cut scores were calculated based on judges' classifications. These two cut scores yielded three proficiency levels. Before they finalized their recommended cut scores, the judges examined how their recommended cut scores affected all New Jersey eighth-grade students who took these tests during the first operational administration in 1999.

The methodology and procedures for the Science standard setting study mirrored those used for the Language Arts Literacy and Mathematics standard setting studies. During the Science standard setting in July 2000, judges examined how their recommended cut scores affected all New Jersey eighth-grade students who took the first operational administration of the Science test in 2000.

5.2 Procedures

Prior to the standard setting studies, descriptions for Proficient and Advanced Proficient performance were developed by independent panels of eighth-grade language arts, mathematics, and science teachers. The proficiency level descriptors were developed to reflect actual test content. Proficiency level descriptors that are anchored in test content allow for more accurate decisions to be made by the judges. The committees developed the following proficiency level descriptors:

Language Arts Literacy–Proficiency Level Descriptors***Proficient***

Eighth-grade students performing at the proficient level are able to construct meaning as they generate their own texts and work with texts generated by others. Proficient students show an overall understanding of the text at literal and inferential levels. They are able to connect with prior knowledge while interacting with, interpreting, and analyzing text.

In reading exercises, students are able to identify and discuss central themes, supporting details, and organizational structures of text. They can extrapolate and synthesize information, monitor their understanding of text, and identify a purpose for reading. Students at this level are able to identify support for and discuss opinions and conclusions as well as to explain textual conventions and literary elements.

Eighth-grade students proficient in their writing are able to develop a central theme, supporting details, and an organizational structure. They establish and sustain a purpose for writing and elaborate on information as they monitor development of text. Students at this level are able to provide support for opinions and conclusions and to use textual and literary elements appropriately.

Advanced Proficient

Eighth-grade students performing at the advanced level are able to construct and extend meaning as they generate their own texts and work with texts generated by others. Advanced students show a sophisticated understanding of abstract themes and ideas that build a text and extend information. They are able to connect with prior knowledge while interacting with, interpreting, analyzing, and critiquing text.

In addition to consistently demonstrating the qualities outlined for a proficient student, the advanced student will demonstrate the ability to synthesize, analyze, and evaluate written text. Students at this level are able to manipulate understanding and will show a high degree of sustained control over textual conventions and literary elements.

Mathematics–Proficiency Level Descriptors

Proficient

The student performing at the proficient level demonstrates evidence of conceptual understanding and of procedural and analytic skills. The student demonstrates the ability to apply mathematical skills and knowledge to theoretical and real-world situations. In addition, the student communicates the required skills and makes connections within and among the mathematical content areas.

The student at this level demonstrates a thorough understanding of basic arithmetic operations—an understanding sufficient for problem solving in practical situations. The student understands the connections between fractions, decimals, percents, and other mathematic topics.

The student understands and applies geometric properties and spatial relationships; applies the principles of similarity, symmetry, and coordinate geometry; interprets data and graphs; determines probabilities; applies the concepts and methods of discrete mathematics, and uses algebraic concepts and processes.

Advanced Proficient

The student performing at the advanced level demonstrates clear and consistent evidence of thorough conceptual understanding, and of procedural and analytic skills. The student consistently demonstrates the qualities outlined for proficient performance. In addition, the student at the advanced level demonstrates the use of abstract thinking and provides explanations that are consistently clear and thorough.

Science–Proficiency Level Descriptors

Proficient

The proficient student can recognize the structural levels of living things. This student knows that some traits of organisms are beneficial and some detrimental. This student can interpret visual and textual data to understand the relationship within a food web and the interdependence of living and nonliving systems.

The proficient student can recognize the effect force has on an object, trace the flow of energy through a system, and use the properties of matter to identify and separate materials. This student can understand different types of energy and use information from data charts to interpret relationships and predict outcomes.

The proficient student can recognize the existence of a relationship between the moon and tides, recognize the different characteristics of the planets in the solar system, and understand the natural forces that change the surface of the Earth, including chemical and physical weathering.

Advanced Proficient

The advanced proficient student can support scientific conclusions with valid contextual and visual data and make predictions based on the interactions of living things. This student is able to use interpretive skills to analyze visual and textual data in order to solve problems dealing with the application of force and energy.

The advanced proficient student understands the difference between types of energy waves and can recognize and apply experimental principles and empirical data.

The advanced proficient student can recognize the nature of the tides' relationship to Earth, Sun, and moon; interpret topographical maps; and identify the steps in the process of weathering and erosion.

Judge Selection Process and Criteria

The standard setting process relied on expert judgments. Therefore, nominations were solicited from school districts for teachers or administrators representing excellence in the teaching profession in terms of knowledge of content area, knowledge of eighth-grade students' skills and abilities, and some understanding of assessment procedures. It was considered critical that these judges represent the more general body of expert New Jersey public school educators. Special care was taken to select judges who were representative of the various District Factor Groups (DFGs) within the state. Additionally, districts were specifically asked to include special education, ESL, and bilingual teachers among their nominees. Districts were also encouraged to nominate members of underrepresented minority populations, e.g., African American or Hispanic, in order to ensure an appropriate diverse representation of statewide populations. Other criteria used in the selection process included number of years teaching experience, the level of content knowledge and student understanding possessed by the nominees, and active participation in content-area professional associations.

Teachers, educators, and content-area experts selected as judges exemplified the required content-area knowledge, teaching experience, and/or understanding of students necessary for an appropriate and comprehensive standard setting study. Each panelist participating in the process represented the knowledge and understanding of his or her peers throughout the course of the process, lending a balance between diverse opinion and consensus.

A concerted effort was made to balance each content-area panel on the basis of county representation, urban representation, representation of schools serving various sizes of populations, gender, and race/ethnicity. The overarching goal of consensus in this forum was not the unanimous agreement of all parties, but the bringing together of individual divergent experiences to form a common understanding of student performance in a content-area that is truly larger, and broader, than its individual parts. The judges selected for the standard setting study represented the same diversity of people and demographics as the students being assessed.

Holistic/Paper Sorting Methodology

The judges' task was to classify student work into one of three performance categories defined to capture levels of performance as expressed by the Partially Proficient, Proficient, and Advanced Proficient categories. The method was holistic in that the judges considered the whole of an individual student's open-ended and multiple-choice responses, i.e., all the items of a particular student for a content area. With the holistic sorting method, the judges reviewed folders of student papers sampled to represent the full range of scores and were asked to sort these folders into three performance levels as represented by the quality of the students' work. An outline of the standard setting procedures follows:

Overview of the 8-Step Plan

Large-Group Session

The standard setting study began with a large-group session. All judges and participants listened to introductory comments and directions for the three-day meeting. The definitions of the standards, their purpose, and ultimate use were discussed. This session was designed to provide a common orientation to judges across content areas.

Step 1 – Description of the Standard Setting Process

Judges worked in their own content area and in separate rooms for the remainder of the process. Step 1 provided the judges with an introduction to the process, their role in the process, and a review of the purpose of the standards.

- *Introductions*
- *Judge Selection Process and Criteria*
- *Purpose of the Standards*
- *Standard Setting Process*
- *Review of the Agenda*
- *Administrative Tasks*

Step 2 – Review of the Assessment Material

Judges became familiar with the assessment at this point. They took the assessment under standardized conditions to get a feel for the experience and content. Judges were also introduced to the content validity evidence for the assessment and the open-ended scoring procedures.

- *Review of Test Content*
- *Brief Description of the Assessment Development Process*
- *Administration of the Assessment to Judges*
- *Scoring the Assessment*

Step 3 – Defining the Standards

Step 3 introduced judges to the definitions of the standards. Judges used exercises to brainstorm student work which typified the definitions for each standard. Judges did not write or re-write the definitions at this time. This step only served to familiarize judges with the definitions, which were previously determined, and to help the judges think about students who are at each standard.

- *Definitions of Student Performance Standards*
- *Interpretation of Proficient Performance*
- *Interpretation of Advanced Proficient Performance*
- *Summary of Student Performance Levels*

Step 4 – Introduction of the Standard Setting Process

Step 4 introduced the specific process to the judges. They practiced reviewing student work and sorting student work into three levels of performance – poor, medium, and high. Judges were provided with information about which multiple-choice items were answered correctly on each sample. In addition, scoring rubrics for the open-ended items were reviewed to facilitate the judgment process for the open-ended items.

- *Description of the Holistic Sorting Method*
- *Summary of the Standard Setting Process*
- *Process Check-off*

Step 5 – Round 1: Holistic Classification of a Wide Range of Student Papers

Judges were instructed in the process of completing the rating sheets. Then, judges were given a set of 33 student papers to classify.

The 33 papers were selected to represent the complete range of test scores for each content area. The raw score distribution for a content area was divided into 11 equal intervals. For each interval, three papers were selected to represent a high score, middle score, and low score within the interval. Judges classified each student work sample as representing an Advanced Proficient, Proficient, or Partially Proficient student by the definitions. Judges recorded their classifications on their rating sheets.

Rating sheets were collected and tabulated with results presented to the judges. Classification frequencies for each paper number were shown to the judges. Judges met in small groups to discuss their classifications. Following the discussions, judges were allowed to make changes to their classifications of the student work on their rating sheets.

- *Distribution of Rating Sheets and Instructions*
- *Classification of Papers (Round 1.1)*
- *Discussion of Judges' Ratings*
- *Review of Classifications (Round 1.2)*

Step 6 – Round 2: Holistic Classification of a Targeted Range of Student Papers

Based on the judges' ratings from Step 5, preliminary cut scores for Advanced Proficient and Proficient were determined using a logistic response model regression of paper scores upon classification decisions. Two papers from each score point at the preliminary cut score and in a range of 5 score points above and below that cut score were selected. Approximately 22 papers were selected to target the borderline between Advanced Proficient and Proficient and approximately 22 papers were selected to target the borderline between Proficient and Partially Proficient.

Judges were then given the 44 student papers targeted at the preliminary cut scores. Judges classified each of these 44 papers as typical of an Advanced Proficient, Proficient, or Proficient/Partially Proficient student by the definitions. Like Step 5, rating sheets were collected and tabulated with results presented to the judges. Classification frequencies for each paper number were shown to the judges. Judges met in small groups to discuss their classifications. Following the discussions, judges were allowed to make changes to their classifications of the student work on their rating sheets before these were collected.

- *Distribution of Rating Sheets and Instructions*
- *Classification of Papers (Round 2.1)*
- *Discussion of Judges' Ratings*
- *Review of Classifications (Round 2.2)*

Step 7 – Review of Impact Data

Judges received reports summarizing their individual ratings and the group cut scores after Step 6. They were provided the statewide performance data to judge the impact of group standards. Judges were allowed, if they desired, to change the raw score value of their cut score according to this new information.

- *Introduction of Individual Judgments and Group Cut Scores*
- *Introduction of Impact Data*
- *Final Standard Determinations*

Step 8 – Evaluation of the Standard Setting Process

Judges were encouraged to rate the process using a five-point scale (five being the highest and one being the lowest). Judges were asked to rate the defining and understanding process of Proficient Performance, Advanced Proficient Performance, and Standard Setting Procedures. Finally, they were asked to rate their confidence in the standard setting results. Additionally, open-ended comments were encouraged.

5.3 Results

Judges were provided with graphical data depicting the impact of the resulting cut scores on the actual score distributions of New Jersey eighth-grade students. In other words, if the Proficient cut score is X and the Advanced Proficient cut score is Y, then A percent of the students would be Partially Proficient, B percent of the students would be Proficient, and C percent of the students would be Advanced Proficient. The data were based on more than 88,000 students for each of the content areas.

Judges had an opportunity to review the implications of their standards in the form of impact data. Judges received cumulative frequency distributions of student scores that allowed them to see the percent and number of students in each category given the standards the judges had set.

Table 5.1 presents the cut scores determined by the judges at each round of the standard setting. The numbers in the table indicate the Proficient/Advanced Proficient cut scores in raw score points. The judges' ratings were quite stable from Round 1.1 to the final recommended cut score. Table 5.2 shows the percentage of students achieving at each proficiency level for the total population with the final cut scores.

The final cut score recommendations shown in Table 5.1 were approved and adopted by the New Jersey State Board of Education.

TABLE 5.1

Proficiency-Level Cut Scores

Cut Scores Proficient/Advanced Proficient	Language Arts Literacy	Mathematics	Science
Total Possible Points	62	56	52
Round 1.1	28.6/45.2	24.4/43.5	24.2/40.1
Round 1.2	28.6/44.7	24.2/43.1	23.7/39.3
Round 2.1	28.2/44.7	24.3/42.8	23.0/39.0
Round 2.2	28.5/45.0	24.5/42.7	24.3/40.2
Final	29.5/44.5	24.0/43.0	24.0/40.0

TABLE 5.2

Percentage of Students Achieving Each Performance Level

	Partially Proficient	Proficient	Advanced Proficient
Language Arts Literacy	24.9%	68.8%	6.3%
Mathematics	40.2%	42.7%	17.0%
Science	26.3%	54.5%	19.2%

CHAPTER 6: SCALING AND EQUATING

6.1 Scaling

The individual student scores are reported as scale scores with a range of 100 to 300. The scores 100 and 300 are a theoretical floor and ceiling and may not actually be observed. The scale score of 250 is the cut score between Proficient students and Advanced Proficient students. The scale score of 200 is the cut score between Proficient students and Partially Proficient students. The score ranges are as follows:

<i>Advanced Proficient</i>	<i>250-300</i>
<i>Proficient</i>	<i>200-249</i>
<i>Partially Proficient</i>	<i>100-199</i>

The Partial Credit Model (PCM) is used for scaling and equating the GEPA operational tests. Masters and Wright (1997) provide this description of the Partial Credit Model:

The Partial Credit Model (PCM) is a unidimensional model for the analysis of responses recorded in two or more ordered categories.... it belongs to the Rasch family of models and so shares the distinguishing characteristics of that family: separable person and item parameters, sufficient statistics, and, hence, conjoint additivity. These features enable "specifically objective" comparisons of persons and items (Rasch, 1977) and allow each set of model parameters to be conditioned out of the estimation procedure for the other.

The PCM (Masters, 1982, 1987, 1988a, 1988b) is the simplest of all item response models for ordered categories. It contains only *two* sets of parameters: one for persons and one for items. All parameters in the model are *locations* on an underlying variable. (p. 101)

WINSTEPS was used to provide the Rasch analyses used for generating the item and student statistics.

Raw score to scale score conversion tables for each content area of regular forms, Braille forms, and breach forms are shown in Appendix D. Appendix E shows Language Arts Literacy, Mathematics, and Science scale score frequency distributions.

6.2 Equating

Equating designs must take into account the form of the assessment. Two equating designs are used. Mathematics and Science are equated using a common anchor item, non-equivalent group, design in which all students take common items. These common items are selected to be representative of the total test form in terms of content, difficulty, and format.

The structure of the Language Arts Literacy does not allow for a subset of common exercises to be selected for use across test administrations because the smallest item exercises are unique and singular.

Reading Comprehension is divided into two passage types. These two types cannot be thought of as representative of each other. The Language Arts Literacy equating is accomplished using an embedded equating/field test section that is used for common-item equating.

Mathematics and Science Equating Design

Common-item equating is used to determine form equivalence from one form, or test administration year, to the next. A set of common (anchor) operational items from the 2005 Mathematics and Science tests was embedded in the 2006 tests. The anchor items include both multiple-choice and open-ended items. Each student participating in the Mathematics and Science testing took the set of common items, and these items contributed to the student's total score. To the maximum extent possible, these items were selected to be proportionally representative of the content and statistics of the total test forms. In addition, the anchor items occupied similar locations in the 2005 and 2006 test forms. These sets of anchor items (14 items with a total of 18 points in Mathematics and 13 items with a total of 15 points in Science) represent approximately one-third of the Mathematics and Science operational tests in terms of number of items and number of points.

The following were applied:

Calibrate the 2006 test items using the Partial Credit Model and fix the item difficulties to their estimated values based on the 2005 calibration. The common set of items is used. The item difficulties for the common anchor items on the spring 2006 test were fixed to the estimated item difficulties from the calibration of the 2005 operational test. This placed all parameter estimates for the 2006 calibration on the 2005 scale. This also produced the new raw score to ability (theta) table for the 2006 test.

Develop a raw score to scale score table for the 2006 assessments. Using the ability to scale score relationship found in the 2005 test calibrations, scale scores were assigned to the raw scores from the 2006 assessments. This was possible because each ability in the ability to scale score table corresponds to a single raw score; therefore, the scale score assigned to that ability can also be assigned to the raw score.

Checks during the equating process were necessary to establish the stability of the common items and determine model fit. One such check was accomplished through the use of the common anchor items from the 2005 operational test embedded in the 2006 operational test. The following is a summary of the steps used for the anchor item analysis.

1. Identify anchor item difficulties from the item bank,
2. Calibrate 2006 form without fixing anchor item difficulties with WINSTEPS,
3. Calculate mean of the bank anchor items difficulties,
4. Calculate mean of 2006 anchor items,
5. Add constant to 2006 anchor item difficulties so the mean equals that found in the bank values,
6. Subtract 2006 and the bank anchor difficulties after adding the constant,
7. Drop item with largest absolute difference greater than or equal to 0.30 for consideration as anchor item, and
8. Repeat steps 1-7 using remaining anchor items.

The final product from the equating procedure was the raw score to scale score table developed in Step 2. When equating was completed, raw score to scale score conversion tables were available for scoring. These two steps can be applied for future assessments.

Language Arts Literacy Equating Design

Scaling and equating for Language Arts Literacy was accomplished through a different design. Each assessment has an embedded equating/field test section that is used for either common-item equating or new-item field testing. Language Arts Literacy was equated using a design in which operational items appeared in a section designated for equating or field testing.

The test included the operational items and four equating sections. Students across the state took one of the equating sections or a field test section. Sampling was done by school and stratified by District Factor Grouping to approximate equivalent groups between equating sets. Sample sizes for each equating/field test form were approximately 8,000 students or more than 8 percent of the student examinee population.

The Language Arts Literacy was equated using a common item design with a combined run. Two forms of the 2006 assessment contained two of the operational passages from 2005 in the field test section. This design allowed for the development of a matrix design in the data, with a combination of data records from 2005 and 2006. All data was analyzed in a combined run with the 2005 item parameters fixed to their 2005 values. This places the 2006 item parameters onto the 2005 scale. Using those 2006 item parameters, a raw score to theta relationship was calculated. This was then used to develop the raw score to scale score table.

Summary of Equating Statistics

Table 6.1 shows a summary of the statistics used to evaluate the psychometric quality of the assessments. All three assessments had a high degree of reliability ranging from 0.87 to 0.90. The standard errors in terms of raw scores ranged from 2.67 to 3.38.

Examination of the fit statistics shows that the Partial Credit Model fits the data reasonably well. The INFIT statistic is a measure of the model fit weighted by the placement of the person locations. The OUTFIT statistic does not apply this weighting and is more sensitive to misfit. It is generally accepted that items with statistics between 0.7 and 1.3 have good fit. On average all assessments demonstrated fit within these limits. On an individual bases, all items had INFIT statistics within this range, but some of the items had OUTFIT statistics which fell outside this range. The number of items with OUTFIT statistics falling outside the range of 0.7 and 1.3 is consisted with past analysis.

Table 6.2 lists the cut scores resulting from the current equating results. Also, those derived from 2002, 2003, 2004, and 2005 are provided for comparison.

TABLE 6.1
Summary of Equating Statistics

	Language Arts Literacy	Mathematics	Science
Number of items	26	36	48
Raw Score Range	0 to 54 by halves	0 to 48 by halves	0 to 54 by halves
Coefficient Alpha	.87	.90	.88
Count of negative biserials	None	None	None
Raw Score (Population)			
Mean	33.0	28.4	30.1
SD	7.4	10.7	9.5
SEM	2.67	3.38	3.29
Rasch Person Measures*			
Mean	1.07	.33	.57
SD	1.29	1.22	.93
SEM	.47	.39	.32
Item Infit MNSQ			
Mean	.94	.99	.99
SD	.47	.23	.17
# Between 0.7 and 1.3	26 of 26	36 of 36	48 of 48
Item Outfit MNSQ			
Mean	1.00	1.01	.99
SD	.68	.33	.21
Between 0.7 and 1.3	20 of 26	31 of 36	46 of 48

TABLE 6.2
Cut Scores and Associated Thetas for Proficiency Levels

	Raw Score Cuts		Rasch Theta Score Cuts	
	Proficient	Advanced	Proficient	Advanced
Language Arts Literacy				
2002	26.5	44.0	0.253	2.780
2003	29.5	45.0	0.238	2.715
2004	31.0	46.5	0.244	2.773
2005	29.0	41.0	0.253	2.664
2006	29.0	41.0	0.223	2.715
Mathematics				
2002	24.0	39.0	-0.074	1.297
2003	24.0	38.5	-0.061	1.323
2004	24.0	38.0	-0.079	1.278
2005	25.0	39.0	-0.062	1.333
2006	25.0	38.5	-0.079	1.278
Science				
2002	22.0	39.5	-0.132	1.344
2003	22.0	39.5	-0.157	1.319
2004	21.0	38.0	-0.132	1.352
2005	20.5	37.5	-0.174	1.344
2006	22.0	39.0	-0.169	1.340

CHAPTER 7: TEST STATISTICS

7.1 Reliability of the Test Scores

Table 7.1 summarizes reliability estimates for the 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 are expressed in terms of the raw score metric and the scale score metric. The scale scores range from 100 to 300.

Reliabilities and SEMs for the dichotomously scored items in each cluster are reported using the Kuder-Richardson Formula 20 (KR-20) in Table 7.2.

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 7.1 and 7.2 reflect the expected positive relationship between test length and reliability.

The SEMs 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 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 scores will lie between $\pm 1Se$ of his true score approximately 68 percent of the time, or $\pm 2Se$ of his true score about 95 percent of the time. Of course, we do not know the true score, but one can infer with about 68% (or 95%) certainty that a person's true score is within $\pm 1Se$ (or $\pm 2Se$) of his observed score. (p. 252)

TABLE 7.1

**Reliability Estimates and Standard Errors of Measurement (SEMs)
for Content Areas and Clusters - 2006**

GEPA Test Section	Number of Points	Reliability Cronbach's alpha	SEM Raw Score	SEM Scale Score
Language Arts Literacy	54	0.88	2.58	11.86
Reading	36	0.88	2.02	–
Writing	18	0.66	1.29	–
Interpreting Text	20	0.79	1.55	–
Analyzing/Critiquing Text	16	0.77	1.30	–
Mathematics	48	0.91	3.25	12.86
Number and Numerical Operations	12	0.72	1.66	–
Geometry and Measurement	12	0.70	1.70	–
Patterns and Algebra	12	0.68	1.59	–
Data Analysis, Probability, and Discrete Mathematics	12	0.70	1.57	–
Knowledge	48	0.91	3.25	–
Problem Solving	36	0.88	2.88	–
Science	54	0.88	3.31	10.89
Life	22	0.78	2.03	–
Physical	16	0.61	1.92	–
Earth	16	0.68	1.79	–
Knowledge	12	0.57	1.54	–
Application	42	0.86	2.93	–

TABLE 7.2

**Reliability Estimates and Standard Errors of Measurement (SEMs)
for Dichotomously Scored Items Within Content Clusters - 2006**

GEPA Content Area	Number of Items	Reliability (KR-20)	SEM Raw Score
Language Arts Literacy	20	0.83	1.66
Reading	20	0.83	1.66
Writing	–	–	–
Writing/Speculate	–	–	–
Writing/Persuade	–	–	–
Interpreting Text	12	0.73	1.34
Analyzing/Critiquing Text	8	0.68	0.98
Mathematics	30	0.86	2.29
Number and Numerical Operations	6	0.57	1.05
Geometry and Measurement	9	0.67	1.29
Patterns and Algebra	9	0.65	1.26
Data Analysis, Probability, and Discrete Mathematics	6	0.59	0.92
Knowledge	30	0.86	2.29
Problem Solving	18	0.80	1.76
Science	45	0.86	2.95
Life	19	0.75	1.89
Physical	13	0.55	1.61
Earth	13	0.64	1.59
Knowledge	12	0.57	1.54
Application	33	0.82	2.51

* There were no dichotomously scored writing items.

CHAPTER 8: ITEM-LEVEL STATISTICS

The GEPA test specifications are aligned with the Core Curriculum Content Standards. Please refer to the *Technical Manual* and Part 2 of this *Technical Report* for information about the test specifications and test development.

8.1 Classical Item Statistics

In Table 8.1, 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 is calculated by dividing students' average score 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 are expressed in terms of the raw score metric.

TABLE 8.1

Item Difficulty and Discrimination Summary Statistics for Dichotomously Scored and Open-Ended Items by Test Section and Cluster - 2006

GEPA Test Section/Cluster	Dichotomous			Open-Ended		
	Item Difficulty		Item Discrimination	Item Difficulty		Item Discrimination
	Mean	S.D.	Mean	Mean	S.D.	Mean
Language Arts Literacy	0.76	0.10	0.47	0.50	0.12	0.93
Reading	0.76	0.10	0.47	0.45	0.14	0.88
Writing	–	–	–	0.55	0.12	0.85
Picture	–	–	–	0.57	0.12	0.74
Persuasive	–	–	–	0.54	0.14	0.81
Interpret Text	0.74	0.11	0.46	0.46	0.15	0.82
Analyze/Critique Text	0.80	0.07	0.48	0.44	0.16	0.83
Mathematics	0.63	0.15	0.44	0.53	0.27	0.94
Number and Numerical Operations	0.63	0.10	0.44	0.52	0.31	0.85
Geometry and Measurement	0.56	0.15	0.43	0.46	0.41	0.73
Patterns and Algebra	0.63	0.15	0.43	0.57	0.35	0.66
Data Analysis	0.75	0.13	0.44	0.55	0.29	0.81
Knowledge	0.63	0.15	0.44	0.53	0.27	0.94
Problem Solving	0.63	0.15	0.44	0.53	0.27	0.94
Science	0.59	0.15	0.37	0.38	0.25	0.83
Life	0.60	0.15	0.39	0.23	0.28	0.66
Physical	0.61	0.15	0.32	0.49	0.38	0.65
Earth	0.57	0.16	0.37	0.43	0.30	0.62
Knowledge	0.58	0.16	0.34	–	–	–
Application	0.60	0.15	0.37	0.38	0.25	0.83

Tables 8.2, 8.3, and 8.4 present frequency distributions of item difficulty (p-values) and item discrimination indices by content cluster. The top section of each table shows the distribution of item difficulty values; the bottom section shows the distribution of point-biserial correlations.

Point-biserial indices are produced to evaluate operational test items. Millman and Greene (1989) note that the point-biserial index gives a true reflection of the item's contribution to the functioning of the test. For field test item review (described in Test Development) biserial correlations are computed. The biserial indices tend to be more stable across samples.

TABLE 8.2

**Frequency Distributions of Item Difficulty
and Item Discrimination by Content Cluster**

2006 Language Arts Literacy

Item Statistics	Interpreting Text	Analyzing Text	Total
ITEM DIFFICULTY: P-VALUES			
.90+	0	0	0
.80 – .89	4	4	8
.70 – .79	5	3	8
.60 – .69	1	1	2
.50 – .59	2	0	2
<.40 – .49	0	0	0
MEAN P-VALUE	.74	.80	.76
MEDIAN P-VALUE	.74	.80	.77
ITEM DISCRIMINATION: POINT-BISERIAL CORRELATIONS			
.50+	4	4	8
.40 – .49	5	3	8
.30 – .39	3	1	4
<.30	0	0	0
MEAN POINT-BISERIAL	.46	.48	.47
MEDIAN POINT-BISERIAL	.45	.50	.46
TOTAL NUMBER OF ITEMS	12	8	20

TABLE 8.3

Frequency Distributions of Item Difficulty and Item Discrimination by Content Cluster
2006 Mathematics

Item Statistics	Number and Numerical Operations	Geometry and Measurement	Patterns and Algebra	Data Analysis, Probability, and Discrete Mathematics	Knowledge	Problem Solving	Test Total
ITEM DIFFICULTY: P-VALUES							
.90+	0	0	0	1	1	1	1
.80 – .89	0	0	1	1	2	1	2
.70 – .79	2	2	3	2	9	5	9
.60 – .69	2	0	1	1	4	3	4
.50 – .59	2	4	0	1	7	4	7
.40 – .49	0	2	4	0	6	3	6
.30 – .39	0	1	0	0	1	1	1
<.30	0	0	0	0	0	0	0
MEAN P-VALUE	.63	.56	.63	.75	.63	.63	.63
MEDIAN P-VALUE	.67	.54	.65	.77	.66	.66	.66
ITEM DISCRIMINATION: POINT-BISERIAL CORRELATIONS							
.50+	1	3	2	2	8	7	8
.40 – .49	3	3	5	2	13	5	13
.30 – .39	2	3	1	1	7	4	7
<.30	0	0	1	1	2	2	2
MEAN POINT-BISERIAL	.44	.43	.43	.44	.44	.44	.44
MEDIAN POINT-BISERIAL	.45	.43	.42	.46	.44	.46	.44
TOTAL NUMBER OF ITEMS	6	9	9	6	30	18	30

TABLE 8.4

**Frequency Distributions of Item Difficulty
and Item Discrimination by Content Cluster**

2006 Science

Item Statistics	Life	Physical	Earth	Knowledge	Application	Total Test
ITEM DIFFICULTY: P-VALUES						
.80 +	1	1	1	1	2	3
.70 – .79	3	3	3	2	7	9
.60 – .69	6	2	1	1	8	9
.50 – .59	4	4	2	4	6	10
.40 – .49	2	2	5	3	6	9
<.40	3	1	1	1	4	5
MEAN P-VALUE	.60	.61	.57	.58	.60	.59
MEDIAN P-VALUE	.61	.57	.56	.55	.61	.60
ITEM DISCRIMINATION: POINT-BISERIAL CORRELATIONS						
.50 +	1	0	0	0	1	1
.40 – .49	10	2	3	3	12	15
.30 – .39	5	7	10	6	16	22
.20 – .29	3	3	0	3	3	6
<.20	0	1	0	0	1	1
MEAN POINT-BISERIAL	.39	.32	.37	.34	.37	.37
MEDIAN POINT-BISERIAL	.40	.32	.35	.33	.39	.37
TOTAL NUMBER OF ITEMS	19	13	13	12	33	45

8.2 Speededness

The amount of time allotted for students to complete the test is intended to provide nearly all students with sufficient time to answer all the questions. Table 8.5 presents data concerning the extent to which this intent was met. Open-ended items appear at the end of each part. For this reason, Table 8.5 shows the percentage of students omitting each of the last three multiple-choice items in each part and all open-ended items.

The percent of students omitting the Reading multiple-choice items is very small, at about 0.2%. The percent of students omitting the open-ended items varies from 1.2% to 5.0%.

TABLE 8.5

**Percentage of Students Omitting the
Last Items of Each Test Part - 2006**

Test Section	Multiple-Choice		Open-Ended	
	Item Number	Percentage Omitting	Item Number	Percentage Omitting
Reading				
Part A	Item 8	0.2%	Item 11	1.2%
	Item 9	0.2%	Item 12	5.0%
	Item 10	0.3%		
Part A	Item 8	0.2%	Item 11	1.8%
	Item 9	0.2%	Item 12	3.2%
	Item 10	0.3%		
Mathematics				
Part A	Item 8	0.1%	Item 11	1.8%
	Item 9	0.8%	Item 12	5.2%
	Item 10	0.3%		
Part B	Item 8	0.5%	Item 11	3.9%
	Item 9	0.2%	Item 12	3.0%
	Item 10	0.3%		
Part C	Item 8	0.2%	Item 11	3.0%
	Item 9	0.2%	Item 12	7.2%
	Item 10	0.4%		
Science				
Part A	Item 13	0.3%	Item 16	4.0%
	Item 14	0.6%		
	Item 15	0.8%		
Part B	Item 13	0.2%	Item 16	2.9%
	Item 14	0.3%		
	Item 15	0.4%		
Part C	Item 13	0.3%	Item 16	3.8%
	Item 14	0.3%		
	Item 15	0.5%		

The percent of students omitting the Mathematics multiple-choice items ranges from 0.1% to 0.8%. The percent of students omitting the Mathematics open-ended items varies from 1.8% to 7.2%.

The percent of students omitting the Science multiple-choice items ranges from 0.3% to 0.8%. The percent of students omitting the Science open-ended items varies from 2.9% to 4.0%.

Overall, these data indicate that the amount of time provided for completing the test is appropriate and that speed of response is not a factor that affects students' performances or detracts from the validity of scores.

8.3 Intercorrelations

The Pearson product-moment correlation between student scores on Language Arts Literacy and Mathematics was .75, Language Arts Literacy and Science was .72, and Mathematics and Science was .81. Table 8.6 shows the correlations between students' scores in the major content clusters and item types. Table 8.7 shows the correlations between student 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 many items that make up a large proportion of the content area score increase the cluster with content area correlation.

For example, the correlation between Reading and Language Arts Literacy in Table 8.6 is quite high (.98) because 36 Reading points are part of the total Language Arts Literacy 54 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. In Table 8.7, the L3 Writing/Speculate cluster has only six points, so this cluster may not correlate as highly with other clusters due to this small number of points.

TABLE 8.6

Intercorrelations Among Major Content Clusters and Item Types - 2006

Major Content and Item Types	Major Content Clusters and Item Types										
	Language Arts Literacy					Mathematics			Science		
	LAT	R	R MC	R OE	W	M T	M MC	M OE	ST	S MC	S OE
LAT Language Arts Literacy (54)											
R Reading (36)	.98										
R MC Reading Multiple-Choice (20)	.92	.96									
R OE Reading Open-ended (16)	.88	.86	.68								
W Writing (18)	.85	.73	.63	.74							
MT Mathematics (48)	.75	.74	.70	.65	.64						
M MC Mathematics Multiple-Choice (30)	.71	.69	.66	.60	.59	.96					
M OE Mathematics Open-ended (18)	.74	.72	.67	.65	.63	.94	.82				
ST Science (54)	.72	.72	.69	.65	.58	.81	.77	.76			
S MC Science Multiple-Choice (45)	.70	.70	.68	.58	.55	.79	.76	.74	.99		
S OE Science Open-ended (9)	.65	.64	.60	.58	.54	.70	.66	.68	.83	.73	

Number in parentheses is the number of score points.

Language Arts Literacy N = 105,437; Mathematics N = 107,419; Science N = 107,489.

TABLE 8.7
Intercorrelations Among Content Areas and Clusters - 2006

Test Section/Cluster	Test Section/Cluster																			
	Language Arts Literacy						Mathematics						Science							
	LAT	L1	L2	L3	L4	L5	L6	MT	M1	M2	M3	M4	M5	M6	ST	S1	S2	S3	S4	S5
LAT Language Arts Literacy (54)																				
L1 Reading (36)	.98																			
L2 Writing (18)	.85	.73																		
L3 Writing/Speculate (6)	.74	.65	.83																	
L4 Writing/Persuade (12)	.81	.69	.97	.66																
L5 Interpreting Text (20)	.94	.96	.69	.61	.65															
L6 Analyzing/Critiquing Text (16)	.93	.94	.70	.62	.66	.82														
MT Mathematics (48)	.75	.74	.64	.55	.61	.72	.69													
M1 Number and Numerical Operations (12)	.68	.66	.58	.50	.55	.64	.62	.91												
M2 Geometry and Measurement (12)	.62	.61	.53	.45	.51	.59	.56	.88	.74											
M3 Patterns and Algebra (12)	.67	.66	.57	.50	.54	.64	.62	.89	.75	.71										
M4 Data Analysis, Probability, and Discrete Mathematics (12)	.72	.71	.60	.52	.57	.68	.67	.89	.75	.70	.74									
M5 Knowledge (48)	.75	.74	.64	.55	.61	.72	.69	1.00	.91	.88	.89	.89								
M6 Problem Solving (36)	.75	.74	.63	.55	.60	.71	.69	.99	.91	.85	.87	.90	.99							
ST Science (54)	.72	.72	.58	.50	.55	.71	.66	.81	.72	.72	.71	.72	.81	.80						
S1 Life (22)	.68	.68	.53	.47	.51	.67	.63	.75	.67	.67	.66	.67	.75	.74	.93					
S2 Physical (16)	.64	.63	.52	.45	.49	.62	.58	.70	.63	.63	.62	.63	.70	.70	.87	.71				
S3 Earth (16)	.61	.61	.49	.43	.47	.60	.56	.71	.64	.64	.62	.62	.71	.70	.88	.73	.66			
S4 Knowledge (12)	.55	.55	.43	.38	.41	.54	.50	.62	.56	.57	.55	.55	.62	.61	.82	.74	.71	.75		
S5 Process Skills (42)	.72	.72	.58	.50	.55	.70	.66	.80	.72	.72	.71	.72	.80	.79	.98	.92	.85	.86	.70	

Number in parentheses is the number of score points.
Language Arts Literacy N = 105,437; Mathematics N = 107,419; Science N = 107,489.

CHAPTER 9: TEST VALIDITY

The validity chapter in the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999, p. 9) begins:

Validity refers to the degree to which evidence and theory support the interpretations of test scores entailed by proposed uses of tests. Validity is, therefore, the most fundamental consideration in developing and evaluating tests. The process of validation involves accumulating evidence to provide a sound scientific basis for the proposed score interpretations. It is the proposed uses that are evaluated, not the test itself. When test scores are used or interpreted in more than one way, each intended interpretation must be validated.

Validity logically begins with an explicit statement of the proposed interpretation of test scores, along with a rationale for the relevance of the interpretation to the proposed use.

The purposes served by the GEPA scores are noted in the following paragraph from page 6 of the manual, *Score Interpretation Manual*:

The GEPA is intended to evaluate the progress students are making in mastering the knowledge and skills required by the end of the eighth grade and in mastering the knowledge and skills they will need to pass the HSPA. The GEPA should serve as a primary indicator for identifying those students who may need instructional intervention. The test should also serve as an indicator for determining which local education programs may need revisions to ensure that instructional programs are aligned with the Core Curriculum Content Standards.

What represents a sufficient collection of evidence in the demonstration of test validity has been the subject of considerable research, thought, and debate in the measurement community over the years. Several different conceptions of validity and approaches to test validation have been proposed, and as a result the field has evolved. In 1995, Messick clarified:

The validity issues of score meaning, relevance, utility, and social consequences are many-faceted and intertwined. They are difficult if not impossible to disentangle, which is why validity has come to be viewed as a unified concept (APA, AERA, & NCME, 1985; Messick, 1989). However, to speak of validity as a unified concept does not imply that validity cannot be usefully differentiated into distinct aspects to underscore issues and nuances that might otherwise be downplayed or overlooked, such as the social consequences of performance assessments or the role of score meaning in applied use. The intent of these distinctions is to provide a means of addressing functional aspects of validity that help disentangle some of the complexities inherent in appraising the appropriateness, meaningfulness, and usefulness of score inferences.

In particular, six distinguishable validity aspects are delineated emphasizing content, substantive, structural, generalizability, external, and consequential aspects of construct validity (Messick, 1994, in press). (pp. 5 and 6)

The fifth edition of the *Standards* (1999) recommends establishing the validity of a test through use of a validity argument. The *Standards* (1999) defines a validity argument as “An explicit scientific justification of the degree to which accumulated evidence and theory support the proposed interpretation(s) of test scores.”

The *Standards* (1999) recognized the following possible sources of validity evidence:

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

The present chapter of this report concerning sources of GEPA validity evidence is organized in sections according to the following traditional validity terms: content and curricular validity, construct validity, criterion-related validity, and consequential validity evidence.

For each of the GEPA content areas, New Jersey educators defined the content and skill test specifications. Content area committees assisted with developing the Directory of Test Specifications and Sample Items which delineate specifications used to create the assessments and to measure student proficiency in the knowledge and skills outlined in the Core Curriculum Content Standards.

Test specifications for the GEPA content areas were designed to align with the Core Curriculum Content Standards. The GEPA Content Committees recommended the emphases and priorities reflected in the number of items for each item type and cluster on the test. The 2006 test specifications are based on the standards adopted in 2002 and 2004.

Curriculum developers and teachers use the specifications, along with curriculum frameworks, the standards themselves, and the score reports, to improve instruction at the district, school, and classroom levels. A number of reports have been designed to assist educators with focusing on pertinent information. Report forms designed to meet specific needs extend the effectiveness of a testing program by making it easier to use test results for educational planning. Chapter 10 of this *Technical Report* includes descriptions and examples of the reports.

Beginning with the 1991 EWT due notice testing, the students' essays also have been returned to the districts for distribution to appropriate district staff members for analysis and use in classroom instruction. A manual, *Cycle II Criterion-Based Holistic Scoring: A Writing Handbook* included with the essays, presents the scoring method and criteria used to evaluate student writing and offers suggestions for using the New Jersey's scoring rubrics and student test data to improve classroom instruction. Teachers are encouraged to review the sample responses in the handbook, the annotations on each of the sample responses, and the features of the respective score scales.

The State Department of Education releases a State Summary Report for each content area tested, which contains summary results at the state, district, and school levels as well as statewide results by District Factor Groups (DFG) and special needs districts. Districts are required to report test results to their boards of education and to the public within 30 days after receiving test reports. Analysis and interpretation of the school and district reports is required by the New Jersey Administrative Code (N.J.A.C. 6A:8-4.3(a), (b)).

Further information about the legal and historical background for the GEPA is available at:

<http://www.state.nj.us/njded/code/current/title6a/chap8.pdf>

<http://www.state.nj.us/njded/assessment/history.shtml>

9.1 Content and Curricular Validity (Evidence Based on Test Content)

Content validity is the most relevant and important source of evidence for the GEPA. The validity of the GEPA scores is based on the alignment of the GEPA to the Core Curriculum Content Standards and the knowledge and skills expected of eighth-grade students.

The Core Curriculum Content Standards were developed by teachers and other educational professionals from New Jersey. The Core Curriculum Content Standards outline what students should know and be able to do at a certain grade level. The questions on the GEPA can contain items/concepts included in the grade eight standards as well as for those standards listed for the prior grades.

The content area committees assisted with developing the Directory of Test Specifications and Sample Items for each of the assessed areas. Attributes of New Jersey educators serving on the committees include:

- strong knowledge of the content area,
- familiarity with New Jersey’s Core Curriculum Content Standards for the specific content area,
- understanding of student’s skills and abilities at the eighth-grade benchmark level,
- some understanding of assessment procedures,
- the ability to work effectively in teams,
- a commitment to educational excellence, and
- sensitivity to students’ needs.

The three content area directories are available online at:

<http://www.njpep.org/assessment/TestSpecs/LangArts/TOC.html>

<http://www.njpep.org/assessment/TestSpecs/MathTestSpec/GEPAMath/MathIndex.html>

<http://www.njpep.org/assessment/TestSpecs/ScienceGEPA/index.html>

Sequential procedures of test specification development through operational test approval described in Chapter 2 of this report ensure the content validity of the tests. The item development teams at Measurement Incorporated begin each item development cycle with a review of the Core Curriculum Content Standards and the three directories of test specifications. Using their years of experience with New Jersey item writing and reviews, item writers understand how to develop multiple-choice and open-ended items that tap the appropriate range of skills. They understand the cognitive complexity required within their content area. Items are designed to assess higher-order or critical thinking skills in varied contexts that are familiar to students. Item content for all items, including the writing-task prompts, is carefully reviewed to ensure that items are free from gender, racial, ethnic, and regional bias.

Prior to field testing, all test items are reviewed by the New Jersey Assessment Content and Sensitivity Review Committees as well as the Office of Evaluation and Assessment staff to ensure that items meet GEPA test specifications including appropriate difficulty and skill requirements. Item approval forms used by the Content Review Committees include two categories that address the cognitive complexity of items:

- match to the test specifications
- appropriate difficulty

The Sensitivity Review Committee reviews to ensure that test questions are not offensive and do not reinforce negative stereotypes, and that test questions appropriately reflect multicultural society. Item approval forms used by the Sensitivity Review Committee require each item to be identified as “Definitely Use” or “Revise and Use With Approval” before the item can be included on a field test.

9.2 Construct Validity (Evidence Based on Response and Evidence Based on Internal Structure)

The glossary of *Standards for Educational and Psychological Testing* (1999) presents this definition of construct validity:

A term used to indicate that the test scores are to be interpreted as indicating the test taker’s standing on the psychological construct measured by the test. A construct is a theoretical variable inferred from multiple types of evidence, which might include the interrelations of the test scores with other variables, internal test structure, observations of response processes, as well as the content of the test. In the current standards, all test scores are viewed as measures of some construct, so the phrase is redundant with validity. The validity argument establishes the construct validity of a test. (p. 174)

Item statistics and intercorrelations provide validity evidence related to internal structure. A large percentage of the GEPA score points for each content area come from open-ended and essay test questions. Beginning with the rangefinding process and continuing through statistical review, many of the responses to these questions are scored, reviewed, and discussed by the Content Review Committees members, the NJDOE Content Coordinators, and the Measurement Incorporated staff. These processes have been repeated annually since 1993. Information obtained from students’ responses to these questions provides insight used for test item acceptance, modification, and rejection as well as for future test item development.

Open-ended questions and essays compose about 63% (34/54) of the Language Arts Literacy points, 38% (18/48) of the Mathematics points, and 17% (9/54) of the Science points. Many open-ended items are field tested each year. During 2006, 28 Reading open-ended items, 6 writing prompts, 21 Mathematics open-ended items, and 11 Science open-ended items were field tested. For each open-ended item, the Measurement Incorporated Project Director prepared a brief summary discussing the types of responses with notes about any issues and concerns. This summary was included with a copy of each item, rubric, sample answer, and rangefinding papers for reference during the statistical review.

For all field test items, Pearson Educational Measurement computed item means, response frequencies, biserial correlations (the field test item with the base test total score), and other descriptive statistics. Content Review Committee members used these statistics, their classroom experiences, and the open-ended responses to discuss and explain the processes they believed students were using to provide the correct and incorrect responses to items. Committee members reviewed for concerns related to ambiguity, irrelevant clues, and inaccuracy. Each item must be classified as “Definitely Use” or “Revise and Use with Approval” before it could appear on an operational test.

In addition, several statistics including item difficulty, item discrimination, and item omits are produced for the operational test and printed in each *Technical Report*. Other operational statistics calculated include Pearson product-moment correlations between students’ scores on the operational test content clusters and item types.

9.3 Criterion-Related Validity (Evidence Based on Relations to Other Variables)

The *Standards for Educational and Psychological Testing* (1974) presents this definition of criterion validity:

Criterion-related validities apply when one wishes to infer from a test score an individual's most probable standing on some other variable called a criterion. Statements of predictive validity indicate the extent to which an individual's future level on the criterion can be predicted from a knowledge of prior test performance; statements of concurrent validity indicate the extent to which the test may be used to estimate an individual's present standing on the criterion. The distinction is important. (p. 26)

Sources of evidence related to concurrent and predictive validity for GEPA score interpretations are linked to the purposes that score report information serves for districts, schools, and teachers. The *Score Interpretation Manual* provides procedures for disseminating score reports and using test score information. A section using reports for student-level evaluation notes:

Further examination of a student's knowledge and skill deficiencies should include the analysis of the student's whole profile. Decisions about appropriate instructional programs should be based on examination of a student's classroom test results, grades, anecdotal records, portfolios, checklists, school-level results, and other measures of performance. (p. 38)

An important purpose of the GEPA is its predictive relationship to the High School Proficiency Assessment (HSPA). A study by Zhao, Robinson, and Guo (2006) provides evidence of the predictive relationship between GEPA scores and HSPA scores. The study considered two cohort samples:

- Cohort 1 (n=37,161) includes students who took the GEPA as eighth graders in 2000 and took the HSPA as eleventh graders in 2003.
- Cohort 2 (n=38,653) includes students who took the GEPA in 2001 and the HSPA in 2004.

Because the GEPA and HSPA programs have no common student identifier, GEPA students' names, gender, and date of birth within school districts were used to match to HSPA students' names, gender, and date of birth to identify students' records to use for the study. The authors noted they matched students within school district to reduce mobility impact and data merge concerns. They did not include Limited English Proficient (LEP) and Special Education (SE) students in the study because these students typically show greater score variation across years.

Zhao, Robinson, and Guo found the correlation coefficient 0.72 for the GEPA and HSPA Language Arts Literacy total scores for both Cohort 1 and Cohort 2. The 0.72 correlation coefficient indicates that the GEPA Language Arts Literacy total score explains 51.8% of the variance in the HSPA Language Arts Literacy total score. Similarly, the correlation coefficient 0.85 was determined for the GEPA and HSPA Mathematics total scores for Cohort 1 and Cohort 2 which indicates 72.3% of the variance in the HSPA mathematics total score is explained by the GEPA mathematics score.

Zhao, Robinson, and Guo calculated the number and percentage of students in the cohorts whose GEPA and HSPA Language Arts Literacy and Mathematics total scores were partially proficient and those students whose GEPA and HSPA Language Arts Literacy and Mathematics total scores were proficient or advanced proficient. In addition to determining the number and percentage for all students in the Cohort 1 and Cohort 2 groups in Language Arts Literacy and Mathematics, the percentages for the Special Needs districts as well as the DFG I and DFG J districts were also calculated.

Language Arts Literacy results are included in Table 9.1 and the Mathematics results are included in Table 9.2.

A possible source of criterion-related validity is the relationship of the GEPA scores to those received on the National Assessment of Educational Progress (NAEP).

The New Jersey assessments and NAEP have several similarities and major differences. The New Jersey assessments and the NAEP are based on content standards and frameworks that are revised or replaced on a regular basis to keep them in line with current instructional practices. Likewise, both the NAEP and New Jersey assessments create test specifications based on their respective frameworks that provide guidelines for developing the test items.

However, the New Jersey assessments and NAEP are distinctly different assessments because of:

- context and purpose,
- content and skills measured,
- item difficulty and formats, and
- method used for setting performance standards (i.e. cut points or achievement levels).

TABLE 9.1
LANGUAGE ARTS LITERACY
Percentages of Students Across GEPA/HSPA Proficiency Levels

All
<p>Cohort 1</p> <p>7.6% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores.</p> <p>2.91% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p> <p>Cohort 2</p> <p>9.25% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores.</p> <p>2.07% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p>
Special Needs
<p>Cohort 1</p> <p>14.9% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores.</p> <p>5.82% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p> <p>Cohort 2</p> <p>19.7% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores.</p> <p>3.91% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p>
DFG I and DFG J
<p>Cohort 1</p> <p>2.94% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores.</p> <p>1.12% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p> <p>Cohort 2</p> <p>3.09% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores.</p> <p>0.75% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p>

TABLE 9.2
MATHEMATICS
Percentages of Students Across GEPA/HSPA Proficiency Levels

All
<p>Cohort 1</p> <p>10.43% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores. 4.75% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p> <p>Cohort 2</p> <p>10.75% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores. 3.79% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p>
Special Needs
<p>Cohort 1</p> <p>10.6% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores. 9.04% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p> <p>Cohort 2</p> <p>12.8% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores. 7.09% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p>
DFG I and DFG J
<p>Cohort 1</p> <p>6.11% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores. 2.07% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p> <p>Cohort 2 DFG I and DFG J</p> <p>5.93% of the students received GEPA Partially Proficient scores and HSPA Proficient or Advanced Proficient scores. 1.70% of the students received GEPA Proficient or Advanced Proficient scores and HSPA Partially Proficient scores.</p>

For these reasons, the New Jersey assessments and the NAEP, even in the same content area, may not yield comparable test results.

New Jersey results for the 2006 NAEP Reading, Mathematics, and Science tests for grade eight students included the following:

- **Reading** - The average scale score was 269 (0 to 500 point scale). About 38 percent of the students scored at or above the NAEP Proficient level, while 20 percent of the students scored at the NAEP Below Basic level. In 2003, 37 percent of the students scored at or above the NAEP Proficient level; and in 2003, 21 percent of the students scored at the Below Basic level on the NAEP.
- **Mathematics** - The average scale score was 284 (0 to 500 point scale). About 36 percent of the students scored at or above the NAEP Proficient level. In 1990, 21 percent of the students scored at or above the NAEP Proficient level; in 1992, 24 percent of the students scored at or above the NAEP Proficient level; and in 2003, 33 percent of the students scored at or above the NAEP Proficient level. In 1990, 42 percent of the students scored at the Below Basic level on the NAEP. In 2006, 26 percent of the students scored at the Below Basic level on NAEP.
- **Science** - The average scale score was 153 (0 to 300 point scale). About 33 percent of the students scored at or above the NAEP Proficient level, while 35 percent of the students scored at the NAEP Below Basic level.

Further information about the NAEP and the New Jersey assessments is available online at <http://www.state.nj.us/njded/assessment/naep/nj.shtml>

9.4 Consequential Validity Evidence (Evidence Based on Consequences of Testing)

Standard 13.1 in Chapter 13: Educational Testing and Assessment in Part 3: “Fairness in Testing,” of the *Standards* (1999) addresses intended and unintended consequences. A very similar standard appears as Standard 15.7 in Chapter 15: Testing in Program Evaluation and Public Policy of Part 3. Standard 13.1 is listed below:

When educational testing programs are mandated by school, district, state, or other authorities, the ways in which test results are intended to be used should be clearly described. It is the responsibility of those who mandate the use of tests to monitor their impact and to identify and minimize potential negative consequences. Consequences resulting from the uses of the test, both intended and unintended, should also be examined by the test user. (p. 145)

Beginning with the EWT due notice testing in 1991, the EWT and GEPA scores have provided districts information to help align their curriculum and instruction with the content and skills tested. The *Score Interpretation Manual* was developed to assist in the analysis and interpretation of GEPA score reports. The manual gives examples of uses of test results, discusses the various test scores, provides information about the appropriate score uses, and cautions against inappropriate score use.

Reports such as the District-Designed Reports were developed to provide districts with tools for organizing data to assist with instructional planning. Students’ score information is arranged on District Design Reports according to a school-developed plan to aggregate their students’ performance. School personnel code students’ answer folders following the school’s plan for grouping and organizing reports. For 2006, 104 districts requested District-Designed Reports for selected groups of students.

The return of student essays for instructional purposes has been an important aspect of Cycle II reporting. The *Cycle II Criterion-Based Holistic Scoring: A Writing Handbook* presents information about the scoring method and criteria used to evaluate student writing. The handbook offers suggestions for using New Jersey’s scoring rubrics and student test data to improve classroom instruction.

A number of materials including the *Cycle II Criterion-Based Holistic Scoring: Mathematics and Science Handbook*, *Cycle II Criterion-Based Holistic Scoring: A Reading Handbook*, *Cycle II Criterion-Based Holistic Scoring: A Writing Handbook*, and the Directory of Test Specifications and Sample Items for each of the GEPA content areas give guidance to teachers and curriculum developers for both instructional improvement and alignment.

Longitudinal graphs from 1999-2006 for Language Arts Literacy and Mathematics and from 2000-2006 for Science are available for the following groups:

- All Students
- Subgroups – General Education, Special Education, Limited English Proficient
- Gender – Female, Male
- Ethnicity – White, Black, Asian, Hispanic
- Economic Status – Economically Disadvantaged, Non-Economically Disadvantaged

The longitudinal graphs for the percent proficient and above by economic status appear in

Figure 9.1 for Language Arts Literacy, Figure 9.2 for Mathematics, and Figure 9.3 for Science. The Language Arts Literacy graphs show that the proficient and above scores hovered between 46.2% and 50.6% for the economically disadvantaged students, and between 78.3% and 83.1% for the non-economically disadvantaged students.

The graphs for Mathematics and Science show generally increasing percents of students with proficient and above scores for the both the economically disadvantaged and non-economically disadvantaged groups. The range of percentages of economically disadvantaged students and non-economically disadvantaged students with proficient and above scores ranged as follows for the 1999-2006 Mathematics administrations and the 2000-2006 Science administrations:

- economically disadvantaged students with proficient and above scores ranged from 25.4% in the 1999 Mathematics test administration to 38.4% in the 2006 test administration;
- non-economically disadvantaged students with proficient and above scores ranged from 64.8% in the 1999 Mathematics test administration to 74.2% in the 2006 test administration;
- economically disadvantaged students with proficient and above scores ranged from 36.9% in the 2000 Science test administration to 57.1% in the 2006 test administration; and
- non-economically disadvantaged students with proficient and above scores ranged from 78.2% in the 2000 Science test administration to 87.2% in the 2006 test administration.

The complete group of longitudinal graphs are available online at:

<http://www.state.nj.us/njded/schools/achievement/2007/gepa/graphs.pdf>

FIGURE 9.1
LANGUAGE ARTS LITERACY
Longitudinal Graph by Economic Status

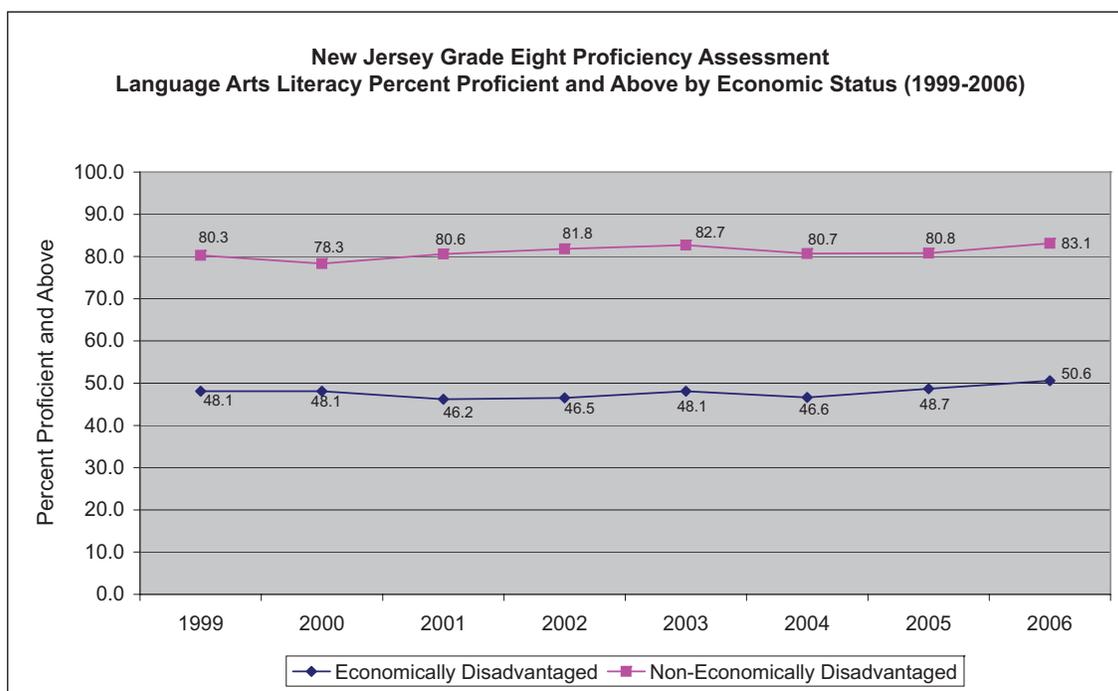


FIGURE 9.2
MATHEMATICS
Longitudinal Graph by Economic Status

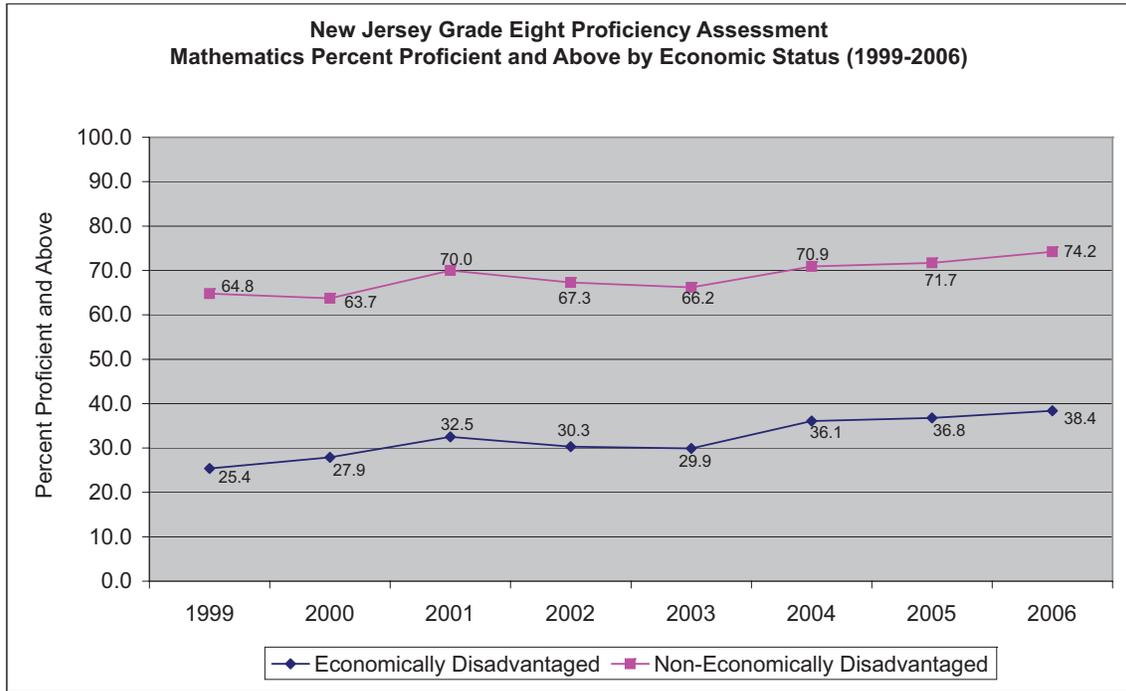
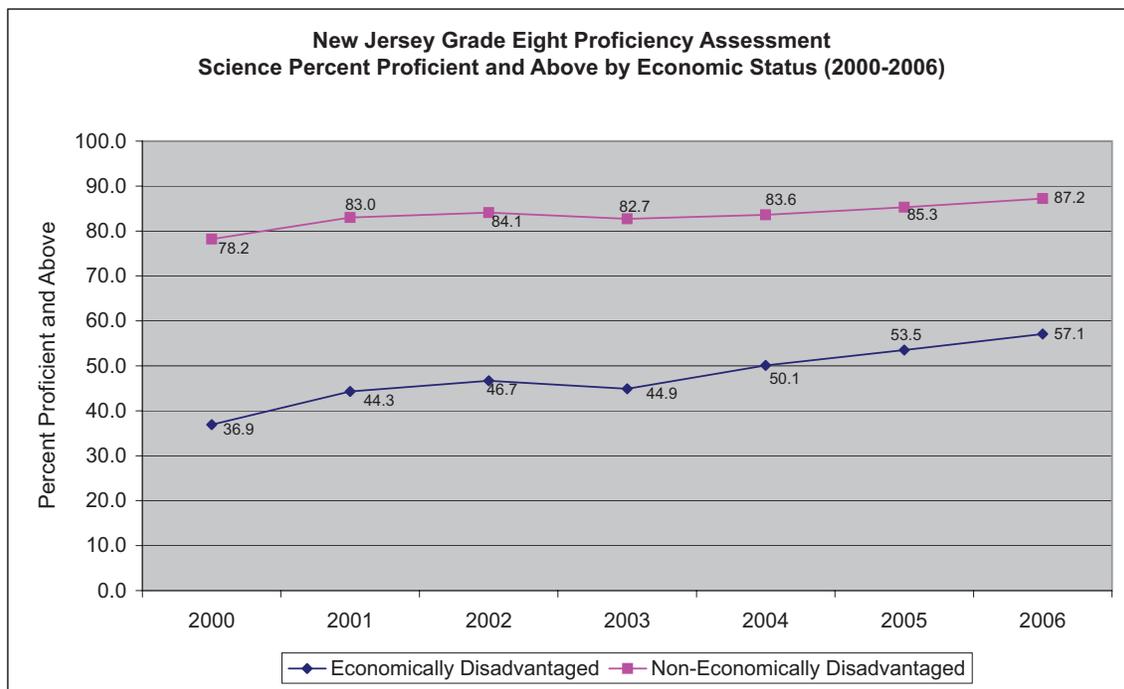


FIGURE 9.3
SCIENCE
Longitudinal Graph by Economic Status



CHAPTER 10: REPORTING

To help school personnel identify the needs of eighth-grade students tested and to assist in the evaluation of school and district programs, a variety of reports are produced and distributed.

The GEPA reports were produced in two cycles:

- Cycle I reports, including Individual Student Reports and preliminary school and district aggregate reports, were received in the districts in mid-June.
- Cycle II reports, including cluster means reports and performance reports for demographic groups, were received in the districts in mid-July.

Cycle II data is used by the Office of Title I Program and Planning and Accountability for Adequate Yearly Progress (AYP) calculations. The State Summary is completed and posted on the NJDOE website in January.

10.1 Information on the Reports

The Cycle I and Cycle II score reports are designed to show a range of student identification and score information to assist school personnel with identifying the needs of their students and recognizing weaknesses in instructional programs.

Student Identification - Score reports display student demographic information gridded on the answer documents or submitted on a pre-ID label files. Prior to reporting, a roster showing the students' demographic information was distributed to school districts to provide an opportunity for corrections.

In addition to the student's name and the Test ID Number assigned to the student, the following information is collected:

- Date of Birth (DOB)
- Gender is indicated by M (male) or F (female).
- Ethnic codes
 - Y (for yes) is indicated in the TIS<1 column if a student was coded as being enrolled in the school for less than a year.
 - Y (for yes) is indicated in the TID<1 column if a student was coded as being enrolled in the district for less than a year.
 - <, 1, 2, 3, or F is indicated in the LEP column if a student was coded as limited English proficient (see LEP in Appendix F). If multiple bubbles were colored, a Y will appear in this column.
- A through N (see SE codes in Appendix F) is indicated in the SE column if a student was coded as a special education student.
- The first letter of a content area (L, M, and S) is indicated in the APA column if a student was coded as taking the APA.
- The first letter of a content area (L, M, and S) is indicated in the T-I column if a student was coded as receiving Title I services for any of the three content areas.
- Y (for yes) is indicated in the ED column if a student was coded as Economically Disadvantaged.
- Y (for yes) is indicated for students coded as having Migrant status.
- Y (for yes) is indicated for students coded by their receiving school [public or private] as being an Out of District placement student.
- Y (for yes) is indicated for students coded as being an Out of Residence Placement student.

Void Codes – Immediately following testing, examiners mark if a student’s answer document should be voided due to illness, disruptive behavior, or some other reason. The answer folder is not scored and a void code is printed in place of the total test score on the student’s reports. These void codes are as follows:

- V1 (voided due to illness)
- V2 (voided due to cheating or disruptive behavior)
- V3 (voided due to the student not being an eighth grader)
- V5 (voided due to breach of security by a school or district).

Also, a student’s answer document may be voided at the time of scoring. For Mathematics and Science, if a student attempted less than 20 percent of the items, no cluster data will appear and, instead of the content area score, the report will list a V4. For Language Arts Literacy, if a student attempted less than 20 percent of the items on one or two testing days but did attempt 20 percent or more on the other testing day, a V4 will appear instead of the Language Arts Literacy score, but cluster data will be provided on the report.

During the 2006 administration, 10 Mathematics and 13 Science tests were voided due to the attempted criteria. For Language Arts Literacy, 176 tests were voided due to the attempted criteria for Day 1 and 262 tests were voided due to the attempted criteria for Day 2.

Score Information – The total GEPA Language Arts Literacy, Mathematics, and Science scores are reported as scale scores with a range of 100 to 300. The scores of 100 and 300 are a theoretical floor and ceiling which may not actually be observed. The scale score of 250 is the cut point between Proficient students and Advanced Proficient students. The scale score of 200 is the cut point between Partially Proficient students and Proficient students. The score ranges are as follows:

<i>Advanced Proficient</i>	<i>250 – 300</i>
<i>Proficient</i>	<i>200 – 249</i>
<i>Partially Proficient</i>	<i>100 – 199</i>

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 and programmatic intervention. District staff should consider multiple measures for all students before making decisions about students’ instructional placement.

In addition to the total GEPA scores in Language Arts Literacy, Mathematics, and Science, various score reports contain the following information for each cluster (scores at the cluster level are raw scores):

- **Points Earned** – This number represents the number of points a student received for a given cluster. On the Student Roster for Language Arts Literacy, the “Points Earned” is provided for Reading and Writing as well as for each of the writing tasks.
- **Just Proficient Mean** – This number represents the average (mean) number of points received for each cluster by all students in the state whose scale scores are 200 for a particular content. Students who took Large-Print or Braille forms are excluded from calculating just proficient means.

Automatic Rescores – The scoring process entails an automatic adjudication of scoring on open-ended items for students whose scores are close to, but not over, the proficiency level. For each content area, the open-ended items of all scale scores ranging from 197 to 199 are automatically rescored to provide the benefit of another examination of student's open-ended scores.

10.2 Types of Reports

Cycle I Reports

Individual Student Report (ISR) and Student Sticker

The Individual Student Report (ISR) is a two-sided report showing specific student score information on the front of the ISR. A description of the GEPA and an interpretation of the ISR scores are printed on the back. Figure 10.1 presents the front of a student's sample report with demographic information, scale scores, proficiency levels, and cluster raw scores and Just Proficient Means. Figure 10.2 shows the GEPA description and ISR interpretation printed for all students.

Two copies of the ISR are produced for every student tested. After educators and school staff analyze the score information on the front of the ISR, one copy is placed in the student's permanent folder and the other copy is shared with the student's parent/guardian in a manner determined by the local district. When a student attends a private school as an Out of District Placement student, a third copy of the ISR is produced and sent to the private school.

A student's scale scores and proficiency levels with the student's identification information are printed on a peel-off label for attaching to a student's permanent folder.

All Sections Roster

The All Sections Roster, an alphabetical listing of students' names, provides students' identification and score information. Each student's scale scores with proficiency levels are listed for the three content areas. Users of this report can quickly determine how a particular student performed in each of the three content areas. The All Sections Roster provides the most complete listing of the student identification information with codes.

FIGURE 10.1

Individual Student Report (ISR Front)



**New Jersey Statewide Assessment System
Grade Eight Proficiency Assessment
Individual Student Report**

Test Date: March 2006	Date of Birth: 01/11/YY	LEP: SE:
County: 99 MIDSTATES	Gender: F	Answer Folder No.: 054321
District: 9999 MIDSTATES		District/School Student ID No:
School: 999 MIDSTATES M.S.		

Student Name: BRADLEY, DENIECE

Content Area	Your Score	Proficiency Level
Language Arts Literacy	235	Proficient
Mathematics	243	Proficient
Science	242	Proficient

Partially Proficient: Score BELOW 200
Proficient: Score AT OR ABOVE 200 but BELOW 250
Advanced Proficient: Score AT OR ABOVE 250

Language Arts Literacy	Mathematics	Science																																																																								
<p>The Language Arts Literacy section assesses a student's abilities in the following clusters. A check mark indicates areas of possible strength.</p> <table border="0"> <tr> <td>Cluster</td> <td>Your Score</td> <td>Just Proficient Mean</td> <td></td> </tr> <tr> <td>Writing</td> <td>15.0 out of 18</td> <td>9.0</td> <td>✓</td> </tr> <tr> <td>Reading</td> <td>25.0 out of 36</td> <td>20.0</td> <td>✓</td> </tr> </table> <hr/> <table border="0"> <tr> <td>Interpreting Text</td> <td>11.5 out of 20</td> <td>11.0</td> <td>✓</td> </tr> <tr> <td>Analyzing/Critiquing Text</td> <td>15.0 out of 16</td> <td>9.0</td> <td>✓</td> </tr> </table>	Cluster	Your Score	Just Proficient Mean		Writing	15.0 out of 18	9.0	✓	Reading	25.0 out of 36	20.0	✓	Interpreting Text	11.5 out of 20	11.0	✓	Analyzing/Critiquing Text	15.0 out of 16	9.0	✓	<p>The Mathematics section assesses a student's abilities in the following clusters. A check mark indicates areas of possible strength.</p> <table border="0"> <tr> <td>Cluster</td> <td>Your Score</td> <td>Just Proficient Mean</td> <td></td> </tr> <tr> <td>Number & Numerical Operations</td> <td>11.0 out of 12</td> <td>5.9</td> <td>✓</td> </tr> <tr> <td>Geometry & Measurement</td> <td>8.5 out of 12</td> <td>5.0</td> <td>✓</td> </tr> <tr> <td>Patterns & Algebra</td> <td>8.5 out of 12</td> <td>6.6</td> <td>✓</td> </tr> <tr> <td>Data Analysis, Probability, & Discrete Mathematics</td> <td>9.0 out of 12</td> <td>7.5</td> <td>✓</td> </tr> </table> <hr/> <table border="0"> <tr> <td>Knowledge</td> <td>37.0 out of 48</td> <td>25.0</td> <td>✓</td> </tr> <tr> <td>Problem Solving Skills</td> <td>18.5 out of 36</td> <td>18.2</td> <td>✓</td> </tr> </table>	Cluster	Your Score	Just Proficient Mean		Number & Numerical Operations	11.0 out of 12	5.9	✓	Geometry & Measurement	8.5 out of 12	5.0	✓	Patterns & Algebra	8.5 out of 12	6.6	✓	Data Analysis, Probability, & Discrete Mathematics	9.0 out of 12	7.5	✓	Knowledge	37.0 out of 48	25.0	✓	Problem Solving Skills	18.5 out of 36	18.2	✓	<p>The Science section assesses a student's abilities in the following clusters. A check mark indicates areas of possible strength.</p> <table border="0"> <tr> <td>Cluster</td> <td>Your Score</td> <td>Just Proficient Mean</td> <td></td> </tr> <tr> <td>Life Science</td> <td>7.0 out of 22</td> <td>8.5</td> <td></td> </tr> <tr> <td>Physical Science</td> <td>15.0 out of 16</td> <td>7.1</td> <td>✓</td> </tr> <tr> <td>Earth Science</td> <td>7.0 out of 16</td> <td>6.4</td> <td>✓</td> </tr> </table> <hr/> <table border="0"> <tr> <td>Knowledge</td> <td>10.0 out of 12</td> <td>5.5</td> <td></td> </tr> <tr> <td>Application</td> <td>19.0 out of 42</td> <td>16.5</td> <td>✓</td> </tr> </table>	Cluster	Your Score	Just Proficient Mean		Life Science	7.0 out of 22	8.5		Physical Science	15.0 out of 16	7.1	✓	Earth Science	7.0 out of 16	6.4	✓	Knowledge	10.0 out of 12	5.5		Application	19.0 out of 42	16.5	✓
Cluster	Your Score	Just Proficient Mean																																																																								
Writing	15.0 out of 18	9.0	✓																																																																							
Reading	25.0 out of 36	20.0	✓																																																																							
Interpreting Text	11.5 out of 20	11.0	✓																																																																							
Analyzing/Critiquing Text	15.0 out of 16	9.0	✓																																																																							
Cluster	Your Score	Just Proficient Mean																																																																								
Number & Numerical Operations	11.0 out of 12	5.9	✓																																																																							
Geometry & Measurement	8.5 out of 12	5.0	✓																																																																							
Patterns & Algebra	8.5 out of 12	6.6	✓																																																																							
Data Analysis, Probability, & Discrete Mathematics	9.0 out of 12	7.5	✓																																																																							
Knowledge	37.0 out of 48	25.0	✓																																																																							
Problem Solving Skills	18.5 out of 36	18.2	✓																																																																							
Cluster	Your Score	Just Proficient Mean																																																																								
Life Science	7.0 out of 22	8.5																																																																								
Physical Science	15.0 out of 16	7.1	✓																																																																							
Earth Science	7.0 out of 16	6.4	✓																																																																							
Knowledge	10.0 out of 12	5.5																																																																								
Application	19.0 out of 42	16.5	✓																																																																							

Note: The scores in this report are for illustrative purposes only. For example, the raw score to scale score conversions are not necessarily the same as those used in the actual reports.

FIGURE 10.2 Individual Student Report (ISR Back)

ABOUT THE GRADE EIGHT PROFICIENCY ASSESSMENT

The Grade Eight Proficiency Assessment (GEPA) was administered in March 2006 to approximately 109,000 eighth-graders throughout New Jersey. The 2006 GEPA measured students' knowledge and skills in three important content areas — Language Arts Literacy, Mathematics, and Science. The GEPA focuses on the knowledge and skills needed for future success in high school and beyond high school.

The GEPA Language Arts Literacy section measures both reading and writing. The Reading component asks students to read real-world texts and to answer related questions. It contains both multiple-choice and open-ended items. The open-ended items require students to write a few sentences or a few paragraphs to answer a question about the text. The Writing component asks students to write two essays. Both the tasks in the Writing component require students to write a response.

The GEPA Mathematics section measures students' abilities to solve problems using mathematical concepts. The components in this section measure: Number and Numerical Operations; Geometry and Measurement. Patterns and Algebra; Data Analysis, Probability, and Discrete Mathematics. The Mathematics section, like the Reading component of the Language Arts Literacy section, contains both multiple-choice and open-ended items. The open-ended items require students to solve a problem as well as explain their solution.

The GEPA Science section measures students' knowledge and skills in Life Science, Physical Science, and Earth Science. The Science section contains both multiple-choice and open-ended items. The open-ended items require students to respond to a question as well as explain their answer.

Carefully trained readers score the open-ended items and essays. Two readers independently score each student's response to ensure the accuracy of each student's results.

The GEPA serves as an early warning for possible difficulties on the High School Proficiency Assessment (HSPA). Beginning with students who entered their junior year in 2001–2002, the HSPA is the "test of record" for receiving a high school diploma. All students are required to receive a score of "proficient" or higher on each section of the HSPA to earn a New Jersey-endorsed high school diploma. (The HSPA replaced the High School Proficiency Test or HSPT.)

HOW TO READ THIS REPORT

This *Individual Student Report* presents your child's Language Arts Literacy, Mathematics, and Science scores on the 2006 GEPA. The report is available only to parents, guardians, students, and authorized school personnel. If you have any questions about the report or how to interpret the scores, you should contact your child's teacher or principal.

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

Additional information to assist in identifying your child's strengths and weaknesses is presented on the bottom half of the report. Cluster-level results show how your child performed on the sets of items that measure particular knowledge and skills (clusters above the line) or particular processes (clusters below the line). Though an item on the GEPA can contribute to a cluster above the line (for example, Reading) as well as a cluster below the line (for example, Interpreting Text), each item is counted only once in the total score.

For each cluster, the column to the right of the cluster name, labeled, **Your Score** shows the number of points your child earned on the clusters. The right-most column for each content area, labeled **Just Proficient Mean**, is a yardstick against which to measure your child's cluster score. It is calculated by taking the average of the raw scores of all the students across the state whose scale scores are 200 for a particular content area. For example, among all students who scored at 200 for the GEPA Language Arts Literacy section, the reading cluster score was 20.0. If your child scored at or above 20.0 for this cluster, a check mark (✓) indicates this area is a possible strength for your child. If your child scored below 20.0 on this cluster, he or she may need additional help in this area.

At the top of the report, a notation may appear if, for some reason, your child's test booklet was not scored. These reasons include illness (V1), cheating or disruptive behavior (V2), or not an eighth grader (V3). For Mathematics and Science, if a student attempted less than 20% of the items, no cluster data will appear and instead of **Your Score** the report will indicate a V4. For Language Arts Literacy, if a student attempted less than 20% of the items on each of the two testing days, no cluster data will appear and instead of **Your Score** for Language Arts Literacy the report will indicate a V4. If a student attempted at least 20% of the items on one of the two testing days, but did not attempt 20% on the other testing day, a V4 will appear in **Your Score** for Language Arts Literacy but cluster scores will be provided on the report. If a student did not take a section of the test, no cluster data will appear and instead of your score, the report will indicate **Not Present**.

Student Roster – Language Arts Literacy, Mathematics, and Science

Three Student Rosters are printed – one for each content area. Students’ names are listed in descending order of the content area scores. Figure 10.3 shows an example of the Student Roster – Mathematics listing the student with the highest score mathematics score first followed with the other students in this school. A dashed line is printed across the roster after the last student in each proficiency level.

No students in the example shown in Figure 10.3 had scores at or above 250, the Advanced Proficient cut point, so a dashed line is printed across the top of the roster. Another dashed line appears across the roster under 200, the Proficient cut point. Students whose answer documents were voided and students who were coded indicating they were taking the Alternate Proficiency Assessment (APA) are listed alphabetically at the end of each content area roster.

Summary of School Performance and Summary of District Performance

A Summary of School Performance is printed for each of the three content areas and a Summary of District Performance is printed for each of the three content areas. The report for each content area provides the number and percent of students in each proficiency level as well as the number of general education students, special education students (including students coded as taking the APA), and limited English proficient students tested for the content area.

The total test information includes the school or district mean for the reported content area. In addition, the means are provided for each of the clusters. The total test and cluster means are printed for the four student groups: total, general education, special education, and limited English proficient.

The following summary information is provided for each subgroup shown on the report:

- Number Enrolled: total number of answer folders returned
- Number Not Present: number of answer folders returned that were totally blank excluding answer folders coded as APA
- Number of Voids: number of answer folders coded void by the school [V1, V2, and V3] AND coded void due to less than 20% of the test items being taken, including answer folders coded as APA [V4] AND coded void due to a security breach [V5]. Number of Valid Score Scores: total number of students tested excluding not present and voids
- Total number of students who scored in each proficiency level, excluding students coded as APA
- Percent of students who scored in each proficiency level, excluding students coded as APA

Preliminary Performance by Demographic Groups – State Report

This preliminary report is produced with the Cycle I reports prior to the completion of the automatic rescoring. The one-page report presents the results for the total, general education, special education, and limited English proficient student groups, and by gender, migrant status, ethnicity, and economic status. Data are based on scale scores and the percentage of students that fall into each of the three proficiency levels. The report does not show cluster level data.

FIGURE 10.3
Student Roster (Mathematics)



New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Student Roster – Mathematics

POINTS EARNED

NUMBER & NUMERICAL OPERATIONS	12 ¹ 6.3 ²	7.0	4.0	5.0	4.0	10.0	25.0	11.0
GEOMETRY & MEASUREMENT	12 ¹ 4.7 ²	7.0	8.5	5.0	4.0	10.0	25.5	9.5
PATTERNS & ALGEBRA	12 ¹ 5.6 ²	4.0	8.0	5.0	4.0	10.0	27.5	10.5
DATA ANALYSIS, PROBABILITY, & DISCRETE MATHEMATICS	12 ¹ 7.4 ²	6.0	6.0	7.5	4.0	10.0	29.0	16.0
KNOWLEDGE	48 ¹ 24.0 ²	11.0	12.0	7.0	3.0	10.0	35.5	14.0
PROBLEM SOLVING SKILLS	36 ¹ 18.3 ²	208	213	218	232	192	186	183

STUDENT NAME/ TEST ID NO.	DATE OF BIRTH	GENDER	LEP	SE	APA	T-I	TOTAL SCORE
GORELOV 0005739748	06/30/YY	F					232
TIPPER I							
DICLEMENTE 0005739705	03/25/YY	M					218
COREY N							
ELLINWOOD 0005739713	10/30/YY	M					213
ELVIS M							
MCGLYNN 0005739519	09/28/YY	F					208
GAIL E							
SPINELLI 0005739845	02/20/YY	M					206
STEVE F							
SCOTT 0005991927	05/16/YY	M					192
RALPH L							
ELLIS 0005991269	08/20/YY	M					186
PERRY M							
BARNES 0005992893	03/23/YY	M					183
KEITH M							
BURKE 0005992923	12/25/YY	M					172
AUSTIN A							
DURBUN 0005991838	05/21/YY	F					189
CASSIE N							
ROSS 0005991552	08/31/YY	F					168
BETSY S							
TAYLOR 0005991561	05/14/YY	M		I			155
JAMES R							
ALSTON 0005992885	03/02/YY	M					V4
ALVIN R							

¹ THE NUMBERS IN THIS ROW ARE THE NUMBER OF POINTS POSSIBLE.
² THE NUMBERS IN THIS ROW ARE THE STATEWIDE RAW SCORE MEANS FOR STUDENTS WHOSE SCALE SCORE IS 200.

Note: The scores in this report are for illustrative purposes only. For example, the raw score to scale score conversions are not necessarily the same as those used in the actual reports.

Preliminary Performance by Demographic Groups – School and District Reports

This report is produced before the rescore is completed. This report does not break the data out at the cluster level. Data are based on scale scores and the percentage of students who fall into each of the three proficiency levels.

Cycle II Reports

The Cycle II reports include a final Performance by Demographic Groups report that reflects any changes that may have occurred during the processing of automatic rescoring.

School and District Cluster Means Reports

Figure 10.4 shows an example of the School Cluster Means Report – Language Arts Literacy. The School and District Cluster Means reports consist of three reports – one for each content area.

The first column on the report presents the mean cluster scores for students in the state whose scale score is 200, i.e., students who are “just proficient.” Data include raw score means of all students (total, general education, special education, limited English proficient, and Title I student groups) at the cluster level for each content area. A similar format is used for both the School Reports and District Reports. The District Reports present aggregated data for the district, DFG, and the state. Additionally, the School Reports show school level data.

District-Designed Reports

The District-Designed Reports are similar to the School Cluster Means Reports except schools create the reports for selected groups of students. Schools used a “special” code category on the GEPA answer documents to obtain cluster means for selected student groups. Like the School Cluster Means Reports, a District-Designed Report is produced for each content area.

Student answer documents may be coded in any of the four two-column “Special Codes” grids labeled A, B, C, or D. These special codes were assigned by the school during the test administrations. The special code, as coded on the students’ answer folders, is printed in the report title. Student groups must contain six or more students.

FIGURE 10.4
Cluster Means Report

**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Cluster Means¹**

Test Date: March 2006
CYCLE II
DISTRICT: 99-Midstate
SCHOOL: 999-Midstate Middle School
Number Enrolled: 429



School

Language Arts Literacy	JUST PROFICIENT MEAN ²	SCHOOL MEAN	DISTRICT MEAN	DFG MEAN	SPECIAL NEEDS MEAN	NON SPECIAL NEEDS MEAN	STATE MEAN
Clusters							
1. Writing (18*)	12.2						
Total Students ³		9.9	9.9	11.2			13.5
General Education		10.5	10.5	12.2			14.9
Special Education		7.3	7.3	7.2			9.4
Limited English Proficient ⁴		8.8	8.8	9.2			9.5
Title I							
2. Reading (36*)	14.3						
Total Students ³		11.9	11.9	14.9			19.5
General Education		12.8	12.8	14.9			11.8
Special Education		8.1	8.1	8.6			11.2
Limited English Proficient ⁴		9.0	9.0	10.3			18.3
Title I							
3. Interpreting Text (20*)	5.9						
Total Students ³		5.2	5.2	5.6			7.6
General Education		5.6	5.6	5.2			8.1
Special Education		3.6	3.6	3.7			5.0
Limited English Proficient ⁴		4.1	4.1	4.3			4.7
Title I							
4. Analyzing/Critiquing Text (16*)	8.4						
Total Students ³		6.7	6.7	8.7			11.6
General Education		7.2	7.2	8.7			5.9
Special Education		4.4	4.4	5.0			6.5
Limited English Proficient ⁴		4.9	4.9	6.0			10.8
Title I							

* Refers to the total raw score points in each cluster.
 1 Cluster means exclude students who took Large print, Braille, and alternate forms as well as students coded as voids.
 2 Just Proficient Means are statewide raw score means for students whose scale score is 200.
 3 Students appear in each applicable category, but they are included in Total Students only once.
 4 Excludes students coded as Former LEP, and students coded as LEP Exempt/LAL.

Performance by Demographic Groups – State Report

Performance by Demographic Groups – State Report summarizes statewide total population data collected from districts regarding general education (GE), special education (SE), LEP, gender, migrant status, ethnicity, and economic status (disadvantaged vs. not disadvantaged). This report includes data from all three content areas. Data are based on scale scores and the percentage of students who fall into each of the three proficiency levels. The report does not break out the data at the cluster level.

The Cycle II Test Results in Appendix B include the Performance by Demographic Groups – State Report.

Performance by Demographic Groups – School and District Reports

Performance by Demographic Groups – School and District Reports present results by general education, special education, LEP, gender, migrant status, ethnicity, and economic status (disadvantaged vs. not disadvantaged) for all three content areas. These group reports provide additional achievement information that can be used to make adjustments to curricula that may better serve these subsections of the total student population. Figure 10.5 shows an example of the school level Performance by Demographic Groups.

Similar to the Performance by Demographic Groups – State Report, data included are based on scale scores and the percentage of students who fall into each of the three proficiency levels. The reports do not break out the data at the cluster level.

FIGURE 10.5

Performance by Demographic Groups

New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
School

TEST DATE: MARCH 2006

REPORT PRINTED

COUNTY: 99 MIDSTATE

DISTRICT: 9999 MIDSTATE

SCHOOL: 999 MIDSTATE M.S.



	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean					
Total Students²	183	4	0	0	179	46.9	45.3	7.8	203.0	4	0	2	177	45.2	44.1	10.7	207.3	4	0	0	179	11.7	61.5	26.8	229.4					
General Education	149	0	0	0	149	37.6	53.0	9.4	209.8	0	0	1	148	35.8	51.4	12.8	213.8	0	0	0	149	4.7	63.1	32.2	235.7					
Special Education	31	4	0	0	27	92.6	7.4	0.0	170.1	4	0	1	26	92.3	7.7	0.0	175.1	4	0	0	27	51.9	48.1	0.0	197.2					
Limited English Proficient ³	4	0	0	0	4	100.0	0.0	0.0	157.0	0	0	0	4	100.0	0.0	0.0	164.3	0	0	0	4	25.0	75.0	0.0	199.0					
Gender⁴																														
Female	82	1	0	0	81	32.1	55.6	12.3	213.5	1	0	0	81	42.0	40.7	17.3	211.3	1	0	0	81	6.2	66.7	27.2	231.6					
Male	101	3	0	0	98	59.2	36.7	4.1	194.3	3	0	2	96	47.9	46.9	5.2	203.9	3	0	0	98	16.3	57.1	26.5	227.5					
Migrant Status																														
Migrant	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0	0.0	0.0	0.0	0.0					
Non-Migrant	183	4	0	0	179	46.9	45.3	7.8	203.0	4	0	2	177	45.2	44.1	10.7	207.3	4	0	0	179	11.7	61.5	26.8	229.4					
Ethnicity																														
White	129	1	0	0	128	39.8	50.8	9.4	206.5	1	0	1	127	41.7	46.5	11.8	209.1	1	0	0	128	8.6	63.3	28.1	232.0					
Black	12	1	0	0	11	54.5	45.5	0.0	196.2	1	0	0	11	54.5	45.5	0.0	201.8	1	0	0	11	9.1	54.5	36.4	230.8					
Asian	8	0	0	0	8	37.5	50.0	12.5	211.9	0	0	0	8	37.5	25.0	37.5	227.1	0	0	0	8	25.0	12.5	62.5	239.5					
Pacific Islander	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0	0.0	0.0	0.0	0.0					
Hispanic	33	1	0	0	32	75.0	21.9	3.1	189.2	1	0	1	31	58.1	38.7	3.2	196.5	1	0	0	32	21.9	68.8	9.4	215.8					
Amer Indian/AK Native	1	1	0	0	0	0.0	0.0	0.0	0.0	1	0	0	0	0.0	0.0	0.0	0.0	1	0	0	0	0.0	0.0	0.0	0.0					
Other ⁵	0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0	0.0	0.0	0.0	0.0	0	0	0	0	0.0	0.0	0.0	0.0					
Economic Status																														
Econ Disadvantaged	32	0	0	0	32	65.6	31.3	3.1	190.9	0	0	0	32	59.4	34.4	6.3	196.8	0	0	0	32	28.1	56.3	15.6	216.9					
Non-Econ Disadvantaged	151	4	0	0	147	42.9	48.3	8.8	205.7	4	0	2	145	42.1	46.2	11.7	209.6	4	0	0	147	8.2	62.6	29.3	232.1					

¹ Includes students coded as LEP Exempt (LAL only)
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.

Note: The scores on this report are for illustrative purposes only.

Appendix A

Scoring Rubrics and 3rd Reader Score Calculation Charts

NEW JERSEY REGISTERED HOLISTIC SCORING RUBRIC

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

NON-SCORABLE RESPONSES	Content/Organization	Usage	Sentence Construction	Mechanics
<p>NR = No Response Student wrote too little to allow a reliable judgment, of his/her writing.</p> <p>OT = Off Topic/ Off Task Student did not write on the assigned topic/task, or the student attempted to copy the prompt.</p> <p>NE = Not English Student wrote in a language other than English.</p> <p>WF = Wrong Format Student refused to write on the topic, or the writing task folder was blank.</p>	<ul style="list-style-type: none"> Communicates intended message to intended audience Relates to topic Opening and closing Focused Logical progression of ideas Transitions Appropriate details and information 	<ul style="list-style-type: none"> Tense formation Subject-verb agreement Pronouns usage/agreement Word choice/meaning Proper Modifiers 	<ul style="list-style-type: none"> Variety of type, structure, and length Correct construction 	<ul style="list-style-type: none"> Spelling Capitalization Punctuation

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

© New Jersey Department of Education

OPEN-ENDED SCORING RUBRIC

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

© New Jersey Department of Education

Holistic Scoring Guide for Mathematics Open-Ended (OE) Items (Generic Rubric)

3-Point Response

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

2-Point Response

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

1-Point Response

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

0-Point Response

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

The above generic rubric is used as a guide to develop specific scoring guides or rubrics for each of the Open-Ended (OE) items which appear on the New Jersey statewide assessments in Mathematics. These scoring rubrics provide the criteria for evaluating and scoring student performance and are developed by a committee of mathematicians and teachers. Rubrics ensure that there is consistency, fairness, and accuracy in scoring open-ended questions.

HOLISTIC SCORING GUIDE (GENERIC RUBRIC) FOR SCIENCE OPEN-ENDED QUESTIONS

The zero-to-three-point generic scoring rubric below was created to help readers score open-ended responses consistently. In scoring, the reader should accept the use of appropriate diagrams, charts, formulas, and/or symbols which are part of a correct answer even when the question does not specifically request their use.

<p>3-Point Response: Student response is reasonably complete, clear, and satisfactory.</p> <p>2-Point Response: Student response has minor omissions and/or some incorrect information.</p> <p>1-Point Response: Student response includes some correct information, but most information included in the response is either incorrect or not relevant.</p> <p>0-Point Response: Student attempts the task but the response is incorrect, not relevant, or inappropriate.</p>

The above generic rubric is used as a guide to develop item specific scoring guides or rubrics for each of the open-ended (OE) questions that appear on the New Jersey statewide assessments in Science. These scoring rubrics provide the criteria for evaluating and scoring student performance and are developed by a committee of scientists and teachers. Rubrics ensure that there is consistency, fairness, and accuracy in scoring open-ended questions.

Table 1

Score Calculation ChartUsed for Means (\bar{x})(Used for 3rd reading equal to or adjacent and all valid)

Absolute Difference ($ 1^{\text{st}} - 2^{\text{nd}} $)	Additional Conditions*	Additional Conditions*	Score Calculation*
0 <i>No 3rd Reading</i>	Both readings are valid	--	$(1^{\text{st}} + 2^{\text{nd}})/2$
1 <i>No 3rd Reading</i>	Both readings are valid	--	$(1^{\text{st}} + 2^{\text{nd}})/2$
2	$1^{\text{st}} < 3^{\text{rd}} < 2^{\text{nd}}$ or $2^{\text{nd}} < 3^{\text{rd}} < 1^{\text{st}}$	--	$(1^{\text{st}} + 2^{\text{nd}})/2$
2	$3^{\text{rd}} < ((1^{\text{st}} + 2^{\text{nd}})/2)$	$1^{\text{st}} < 2^{\text{nd}}$	$(1^{\text{st}} + 3^{\text{rd}})/2$
		$2^{\text{nd}} < 1^{\text{st}}$	$(2^{\text{nd}} + 3^{\text{rd}})/2$
	$3^{\text{rd}} > ((1^{\text{st}} + 2^{\text{nd}})/2)$	$1^{\text{st}} < 2^{\text{nd}}$	$(2^{\text{nd}} + 3^{\text{rd}})/2$
		$2^{\text{nd}} < 1^{\text{st}}$	$(1^{\text{st}} + 3^{\text{rd}})/2$
3	$3^{\text{rd}} = 1^{\text{st}}$ or $(3^{\text{rd}} \pm 1) = 1^{\text{st}}$	--	$(1^{\text{st}} + 3^{\text{rd}})/2$
	$3^{\text{rd}} = 2^{\text{nd}}$ or $(3^{\text{rd}} \pm 1) = 2^{\text{nd}}$	--	$(2^{\text{nd}} + 3^{\text{rd}})/2$
4 and 5	$3^{\text{rd}} = 1^{\text{st}}$ or $(3^{\text{rd}} \pm 1) = 1^{\text{st}}$	--	$(1^{\text{st}} + 3^{\text{rd}})/2$
	$3^{\text{rd}} = 2^{\text{nd}}$ or $(3^{\text{rd}} \pm 1) = 2^{\text{nd}}$	--	$(2^{\text{nd}} + 3^{\text{rd}})/2$

If both readings are invalid and equal, the score is 0.

Used for Means (χ)
(Used for 3rd Reading NOT equal to or adjacent but all valid)

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

Table 2A
Used for Means (χ)
(Used for 1st or 2nd reading invalid and 3rd Reading valid)

Condition	Additional Condition	Score Calculation
1 st Reading Invalid 2 nd Reading Valid	Absolute difference between 2 nd Reading and 3 rd reading is 0 or 1	$(2^{\text{nd}} + 3^{\text{rd}}) / 2$
	Absolute difference between 2 nd Reading and 3 rd reading is greater than 1	Use 3 rd Reading
1 st Reading Valid 2 nd Reading Invalid	Absolute difference between 1 st Reading and 3 rd reading is 0 or 1	$(1^{\text{st}} + 3^{\text{rd}}) / 2$
	Absolute difference between 1 st Reading and 3 rd reading is greater than 1	Use 3 rd Reading
Both 1 st and 2 nd Readings are invalid		Use 3 rd Reading

If the 3rd Reading is invalid, use the 3rd reading score.

Table 3
Score Calculation Chart

Used for Sum (Σ)
 (Used for 3rd Reading equal to or adjacent and all valid)

Absolute Difference (1 st - 2 nd)	Additional Conditions*	Additional Conditions*	Score Calculation*
0 <i>No 3rd Reading</i>	--	--	(1 st + 2 nd)
1 <i>No 3rd Reading</i>	--	--	(1 st + 2 nd)
2 – 5	Equal to or Adjacent	--	$((1^{\text{st}} + 2^{\text{nd}} + 3^{\text{rd}}) * 2) / 3$

Table 4**Additional Score Calculations**Used for Sum (Σ)(Used for 3rd Reading NOT equal to or adjacent but all valid)

Condition	Score Calculation
NOT Equal to or Adjacent	$((1^{\text{st}} + 2^{\text{nd}} + 3^{\text{rd}}) * 2) / 3$

If the 3rd Reading is invalid, use the 3rd reading score.**Table 4A**Used for Sum (Σ)(Used for 1st or 2nd reading invalid and 3rd Reading valid)

Condition	Additional Condition	Score Calculation
1 st Reading Invalid 2 nd Reading Valid	Absolute difference between 2 nd Reading and 3 rd reading is 0 or 1	$(2^{\text{nd}} + 3^{\text{rd}})$
	Absolute difference between 2 nd Reading and 3 rd reading is greater than 1	$(3^{\text{rd}} * 2)$
1 st Reading Valid 2 nd Reading Invalid	Absolute difference between 1 st Reading and 3 rd reading is 0 or 1	$(1^{\text{st}} + 3^{\text{rd}})$
	Absolute difference between 1 st Reading and 3 rd reading is greater than 1	$(3^{\text{rd}} * 2)$
Both 1 st and 2 nd Readings are invalid		Use 3 rd Reading

If the 3rd Reading is invalid, use the 3rd reading score.

Table 5
When to Use the Mean vs. Sum Scoring Rules

Subject	Valid scores	Grade 8
Reading OE	0-4 *	Mean
Writing – Picture	Grade 8 1-6 **	Mean
Writing – Persuasive	1-6 **	Sum
Math OE	0-3 *	Mean
Science OE	0-3 *	Mean

Designation Codes:

- * = 7 = NR, for No Response
(blank, fragmented, refusing or unable to write on topic, copy of item)
8 = OT, for Off Topic
9 = NE, for Not English
- ** = 0 = NR, for No Response
7 = WF, for Wrong Format
8 = OT, for Off Topic
9 = NE, for Not English
- *** = 7 = NR, No Response

There are three types of situations that will require a third reading:

1. First and second reading are valid scores and not equal or adjacent.
2. One reading is a valid score and the other reading is not a valid score.
3. Both readings are not a valid score and are not equal.

Appendix B

Cycle II Test Results

Executive Summary

The 2006 New Jersey Grade Eight Proficiency Assessment (GEPA) consisted of three content areas: Language Arts Literacy, Mathematics, and Science. The GEPA is used as a primary indicator for identifying those students who may need instructional intervention in the three content areas. It is designed to give an indication of the progress students are making in mastering the skills they will need to pass the High School Proficiency Assessment (HSPA).

The GEPA Language Arts Literacy, Mathematics, and Science scores are reported as scale scores with a range of 100 to 300. Please note that 100 and 300 are a theoretical floor and ceiling, which may not actually be observed. The score ranges are as follows:

<i>Advanced Proficient</i>	<i>250-300</i>
<i>Proficient</i>	<i>200-249</i>
<i>Partially Proficient</i>	<i>100-199</i>

Students who scored Partially Proficient are considered to be below the state minimum level of proficiency. These students may need additional instructional support such as individual or programmatic intervention. It is important that districts consider multiple measures with all students before making final decisions about students' instructional placement.

The GEPA was administered between March 13 and March 16, 2006. Of the 109,091 students enrolled, 106,447 students received valid scale scores in Language Arts Literacy, 107,530 students received valid scale scores in Mathematics, and 107,600 students received valid scale scores in Science.

For the total group of students, 25.7% scored Partially Proficient, 65.7% Proficient, and 8.5% Advanced Proficient in Language Arts Literacy. In Mathematics, 35.5% scored Partially Proficient, 43.6% Proficient, and 20.9% Advanced Proficient. In Science, 21.0% scored Partially Proficient, 58.5% Proficient, and 20.5% Advanced Proficient. The mean scale score was 214.3 in Language Arts Literacy, 214.6 in Mathematics, and 223.2 in Science.

This executive summary includes four tables summarizing statewide performance by demographic groups. Tables 1–3 present the performance in Language Arts Literacy, Mathematics, and Science, respectively. Table 4 presents the performance for the state, Special Needs districts, and Non-Special Needs districts.

The performance data include only students with valid scale scores. Students whose answer folders were voided are excluded. Students may receive a scale score in one content area, but not in others.

Beginning in 2006, results for the general education group are not reported in the state summary. Previously, this group included students with answer folders not coded as special education or limited English proficient. A major change for the 2006 State Summary is that Limited English Proficient (LEP) is reported as LEP (Current plus Former) with two subcategories: LEP Current and LEP Former.

Student performance is summarized by total students, education program, and student demographic subgroups: Total, Special Education (SE), Limited English Proficient status (LEP), Gender, Ethnicity, Economic status (disadvantaged vs. not disadvantaged), and Migrant status.

For each demographic group, the number of students participating, the percent of students in each proficiency level, and the mean scale score are reported for each content area. The percentages of students for the three proficiency levels may not total to one hundred due to rounding. The percentage of students in Proficient or Advanced Proficient is calculated by subtracting the percentage of students in Partially Proficient from one hundred.

Demographic information originates from the data collected on the students' answer folders. School district personnel were given an opportunity to review the demographic information they provided on the answer folders and correct any errors prior to reporting.

This executive summary includes information from the state level Performance by Demographic Groups Report from Cycle II reporting. The complete state summary data file with District Factor Groups and longitudinal data is available at <http://www.state.nj.us/njded/schools/achievement/>.

Reporting Rules for State Summary Data File

The state summary data files contain the same type of information shown on the Statewide Performance by Demographic Groups Report for schools and districts included with the Cycle II reporting. In order to safeguard student confidentiality, certain information is suppressed in the files according to the following reporting rules:

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

Highlights from the 2006 GEPA Performance Results

The percentages of students scoring at Proficient or Advanced Proficient by content areas are described below:

Total

- Language Arts Literacy – 74.3% of the students scored Proficient or Advanced Proficient and 8.5% of the students scored Advanced Proficient
- Mathematics – 64.5% of the students scored Proficient or Advanced Proficient and 20.9% of the students scored Advanced Proficient
- Science – 79% of the students scored Proficient or Advanced Proficient and 20.5% of the students scored Advanced Proficient

Special Education

- Language Arts Literacy – 32.7% of the students scored Proficient or Advanced Proficient and 0.5% of the students scored Advanced Proficient
- Mathematics – 25.0% of the students scored Proficient or Advanced Proficient and 3.1% of the students scored Advanced Proficient
- Science – 49.8% of the students scored Proficient or Advanced Proficient and 4.9% of the students scored Advanced Proficient

Limited English Proficient (LEP)

- Language Arts Literacy – 23.9% of the LEP Current and Former students scored Proficient or Advanced Proficient and 0.5% of the LEP Current and Former students scored Advanced Proficient. About 15.8% of the LEP Current students scored Proficient or Advanced Proficient and 0.2% of the LEP Current students scored Advanced Proficient and about 48.4% of the LEP Former students scored Proficient or Advanced Proficient and 1.2% of the LEP Former students scored Advanced Proficient.
- Mathematics – 27.6% of the LEP Current and Former students scored Proficient or Advanced Proficient and 5.8% of the group scored Advanced Proficient. Of the LEP Current students, 22.6% scored Proficient or Advanced Proficient and 4.7% scored Advanced Proficient. Of the LEP Former students, 46.7% scored Proficient or Advanced Proficient and 10.2% scored Advanced Proficient.
- Science – 33.4% of the LEP Current and Former students scored Proficient or Advanced Proficient and 2.5% of the group scored Advanced Proficient. Of the LEP Current students, 26.3% scored Proficient or Advanced Proficient and 1.4% scored Advanced Proficient. Of the LEP Former students, 61.1% scored Proficient or Advanced Proficient and 6.6% scored Advanced Proficient.

Gender

- Language Arts Literacy – 80.3% of the female students and 68.7% of the male students scored Proficient or Advanced Proficient while 11.9% of the female students and 5.3% of the male students scored Advanced Proficient
- Mathematics – 64.1% of the female students and 64.8% of the male students scored Proficient or Advanced Proficient while 18.8% of the female students and 22.9% of the male students scored Advanced Proficient
- Science – 78.1% of the female students and 79.8% of the male students scored Proficient or Advanced Proficient while 17.1% of the female students and 23.7% of the male students scored Advanced Proficient

Ethnicity

- Language Arts Literacy – percentages of Proficient and Advanced Proficient ranged from 85.5% of Asian students and 85.4% of White students to 50.1% of Black students while the percentages of Advanced Proficient ranged from 18.8% of Asian students to 2.1% of Black students and 2.6% of Hispanic students. (The percentages of the Proficient and Advanced Proficient scores in the other ethnic groups fell between the Asian and Black groups.)
- Mathematics – percentages of Proficient and Advanced Proficient ranged from 85.6% of Asian students to 31.7% of Black students while percentages of Advanced Proficient ranged from 44.8% of Asian students to 4.3% of Black students.
- Science – percentages of Proficient and Advanced Proficient ranged from 91.4% of Pacific Islander, 90.3% of White students, and 90.0% of Asian students to 54.7% of Black students while percentages of Advanced Proficient ranged from 36.7% of Asian students to 4.6% of Black students.

TABLE 1
2006 Grade Eight Proficiency Assessment
Statewide Performance
Language Arts Literacy

	Number of Students Enrolled	Number of APA Students	Number Not Present	Number of Voids	Number of Students with Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean
Total Students	109,091	689	604	1,351	106,447	25.7	65.7	8.5	214.3
Special Education	18,327	689	216	346	17,076	67.3	32.2	0.5	184.2
LEP (current and former)	4,007	11	45	708	3,243	76.1	23.4	0.5	177.2
LEP current	3,188	7	43	701	2,437	84.2	15.6	0.2	170.7
LEP former	819	4	2	7	806	51.6	47.1	1.2	196.7
Gender									
Female	52,729	243	255	570	51,661	19.7	68.3	11.9	219.7
Male	56,319	444	342	774	54,759	31.3	63.3	5.3	209.2
Ethnicity									
White	62,478	381	194	306	61,597	14.6	74.3	11.1	222.4
Black	19,563	125	217	304	18,917	49.9	48.0	2.1	196.4
Asian	7,312	36	24	110	7,142	14.5	66.6	18.8	226.3
Pacific Islander	305	0	1	0	304	15.1	75.3	9.5	220.2
Hispanic	18,925	131	156	617	18,021	42.7	54.7	2.6	200.7
American Indian/Alaskan Native	121	1	0	4	116	38.8	55.2	6.0	205.1
Other	387	15	12	10	350	31.4	58.9	9.7	211.0
Economic Status									
Economically Disadvantaged	30,006	227	293	746	28,740	49.4	48.8	1.8	196.6
Non-Economically Disadvantaged	79,085	462	311	605	77,707	16.9	72.0	11.0	220.9
Migrant Status									
Migrant	38	0	0	2	36	66.7	30.6	2.8	189.1
Non-Migrant	109,053	689	604	1,349	106,411	25.7	65.8	8.5	214.3

TABLE 2
2006 Grade Eight Proficiency Assessment
Statewide Performance
Mathematics

	Number of Students Enrolled	Number of APA Students	Number Not Present	Number of Voids	Number of Students with Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean
Total Students	109,091	697	685	179	107,530	35.5	43.6	20.9	214.6
Special Education	18,327	697	244	107	17,279	75.0	21.9	3.1	182.9
LEP (current and former)	4,007	11	27	9	3,960	72.4	21.8	5.8	185.0
LEP current	3,188	7	26	6	3,149	77.4	18.0	4.7	180.8
LEP former	819	4	1	3	811	53.3	36.5	10.2	201.1
Gender									
Female	52,729	249	290	56	52,134	35.9	45.3	18.8	213.7
Male	56,319	446	388	120	55,365	35.2	41.9	22.9	215.5
Ethnicity									
White	62,478	382	241	86	61,769	22.1	50.7	27.2	224.9
Black	19,563	127	256	52	19,128	68.3	27.4	4.3	188.8
Asian	7,312	38	12	4	7,258	14.4	40.8	44.8	237.3
Pacific Islander	305	0	1	0	304	24.3	51.0	24.7	224.0
Hispanic	18,925	134	165	34	18,592	54.5	37.7	7.8	198.2
American Indian/Alaskan Native	121	1	0	0	120	46.7	42.5	10.8	203.6
Other	387	15	10	3	359	43.5	38.2	18.4	208.5
Economic Status									
Economically Disadvantaged	30,006	229	332	69	29,376	61.6	32.2	6.3	193.4
Non-Economically Disadvantaged	79,085	468	353	110	78,154	25.8	47.9	26.4	222.6
Migrant Status									
Migrant	38	0	0	0	38	65.8	28.9	5.3	187.8
Non-Migrant	109,053	697	685	179	107,492	35.5	43.6	20.9	214.6

Economic Status

- Language Arts Literacy – 50.6% of Economically Disadvantaged students and 83.1% of Non-Economically Disadvantaged students scored Proficient or Advanced Proficient while 1.8% of Economically Disadvantaged students and 11.0% of Non-Economically Disadvantaged students scored Advanced Proficient.
- Mathematics – 38.4% of Economically Disadvantaged students and 74.2% of Non-Economically Disadvantaged students scored Proficient or Advanced Proficient while 6.3% of Economically Disadvantaged students and 26.4% of Non-Economically Disadvantaged students scored Advanced Proficient.
- Science – 57.1% of the Economically Disadvantaged students and 87.2% of Non-Economically Disadvantaged students scored Proficient or Advanced Proficient while 5.2% of Economically Disadvantaged students and 26.2% of Non-Economically Disadvantaged students scored Advanced Proficient.

Migrant

Only 0.035% of the enrolled grade 8 students were migrant students. The percentage of Migrant students scoring at Proficient or Advanced Proficient was 33.3% for Language Arts Literacy, 34.2% for Mathematics, and 62.2% for Science. The percentage of Migrant students scoring at Advanced Proficient was 2.8% for Language Arts Literacy, 5.3% for Mathematics, and 2.7% for Science.

TABLE 3
2006 Grade Eight Proficiency Assessment
Statewide Performance
Science

	Number of Students Enrolled	Number of APA Students	Number Not Present	Number of Voids	Number of Students with Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean
Total Students	109,091	665	690	136	107,600	21.0	58.5	20.5	223.2
Special Education	18,327	665	253	76	17,333	50.2	44.9	4.9	200.6
LEP (current and former)	4,007	10	24	3	3,970	66.6	30.9	2.5	191.3
LEP current	3,188	6	24	2	3,156	73.7	24.9	1.4	187.2
LEP former	819	4	0	1	814	38.9	54.4	6.6	206.9
Gender									
Female	52,729	239	289	45	52,156	21.9	61.0	17.1	221.2
Male	56,319	424	395	85	55,415	20.2	56.2	23.7	225.1
Ethnicity									
White	62,478	363	244	54	61,817	9.7	62.7	27.5	232.4
Black	19,563	121	257	50	19,135	45.3	50.1	4.6	202.9
Asian	7,312	37	13	2	7,260	10.0	53.4	36.7	236.7
Pacific Islander	305	0	1	1	303	8.6	65.3	26.1	231.3
Hispanic	18,925	128	164	23	18,610	38.0	55.1	6.9	208.0
American Indian/Alaskan Native	121	1	0	0	120	28.3	57.5	14.2	216.5
Other	387	15	11	6	355	23.7	57.2	19.2	220.7
Economic Status									
Economically Disadvantaged	30,006	217	333	47	29,409	42.9	51.9	5.2	204.5
Non-Economically Disadvantaged	79,085	448	357	89	78,191	12.8	61.0	26.2	230.2
Migrant Status									
Migrant	38	0	1	0	37	37.8	59.5	2.7	199.6
Non-Migrant	109,053	665	689	136	107,563	21.0	58.5	20.5	223.2

Table 4 presents the number of students with valid scale scores and the percentage of students in each proficiency level for the state, Special Needs districts, and Non-Special Needs districts.

Statewide Total Students. The percentage scoring at or above Proficient in each content area was:

- Language Arts Literacy – 74.3% of the 106,447 students with valid scores
- Mathematics – 64.5% of the 107,530 students with valid scores
- Science – 79% of the 107,600 students with valid scores

Total Students in Non-Special Needs districts. The percentage scoring at or above Proficient in each content area was:

- Language Arts Literacy – 80.7% of the 86,106 students with valid scores
- Mathematics – 71.6% of the 86,632 students with valid scores
- Science – 85.5% of the 86,691 students with valid scores

Total Students in Special Needs districts. The percentage scoring at or above Proficient in each content area was:

- Language Arts Literacy – 47.3% of the 20,341 students with valid scores
- Mathematics – 34.9% of the 20,898 students with valid scores
- Science – 51.8% of the 20,909 students with valid scores

TABLE 4
2006 Grade Eight Proficiency Assessment
Statewide Performance
Non-Special Needs and Special Needs Districts

	Number of Students with Valid Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean
LANGUAGE ARTS LITERACY					
STATEWIDE TOTAL	106,447	25.7	65.7	8.5	214.3
Non-Special Needs	86,106	19.3	70.5	10.1	219.0
Special Needs	20,341	52.7	45.5	1.8	194.4
SPECIAL EDUCATION	17,076	67.3	32.2	0.5	184.2
Non-Special Needs	13,024	60.7	38.6	0.7	189.3
Special Needs	4,052	88.4	11.5	0.1	167.7
LEP CURRENT & FORMER	3,243	76.1	23.4	0.5	177.2
Non-Special Needs	1,478	72.9	26.4	0.7	180.0
Special Needs	1,765	78.8	20.9	0.3	174.8
LEP CURRENT	2,437	84.2	15.6	0.2	170.7
Non-Special Needs	1,146	81.7	17.9	0.4	173.7
Special Needs	1,291	86.4	13.5	0.1	168.0
LEP FORMER	806	51.6	47.1	1.2	196.7
Non-Special Needs	332	42.8	55.7	1.5	201.5
Special Needs	474	57.8	41.1	1.1	193.4
MATHEMATICS					
STATEWIDE TOTAL	107,530	35.5	43.6	20.9	214.6
Non-Special Needs	86,632	28.4	47.1	24.5	220.4
Special Needs	20,898	65.1	28.8	6.1	190.8
SPECIAL EDUCATION	17,279	75.0	21.9	3.1	182.9
Non-Special Needs	13,129	70.2	26.0	3.9	187.2
Special Needs	4,150	90.1	9.3	0.7	169.3
LEP CURRENT & FORMER	3,960	72.4	21.8	5.8	185.0
Non-Special Needs	1,805	66.6	24.2	9.2	191.0
Special Needs	2,155	77.3	19.7	3.0	180.0
LEP CURRENT	3,149	77.4	18.0	4.7	180.8
Non-Special Needs	1,471	71.4	21.2	7.4	186.6
Special Needs	1,678	82.6	15.1	2.3	175.8
LEP FORMER	811	53.3	36.5	10.2	201.1
Non-Special Needs	334	45.5	37.4	17.1	210.0
Special Needs	477	58.7	35.8	5.5	194.9
SCIENCE					
STATEWIDE TOTAL	107,600	21.0	58.5	20.5	223.2
Non-Special Needs	86,691	14.5	61.2	24.3	228.4
Special Needs	20,909	48.2	47.4	4.5	201.6
SPECIAL EDUCATION	17,333	50.2	44.9	4.9	200.6
Non-Special Needs	13,178	41.9	51.9	6.2	205.5
Special Needs	4,155	76.7	22.7	0.6	185.2
LEP CURRENT & FORMER	3,970	66.6	30.9	2.5	191.3
Non-Special Needs	1,809	60.6	35.5	3.9	195.4
Special Needs	2,161	71.6	27.1	1.3	187.8
LEP CURRENT	3,156	73.7	24.9	1.4	187.2
Non-Special Needs	1,473	67.9	29.9	2.2	191.1
Special Needs	1,683	78.8	20.5	0.7	183.8
LEP FORMER	814	38.9	54.4	6.6	206.9
Non-Special Needs	336	28.9	60.1	11.0	214.2
Special Needs	478	46.0	50.4	3.6	201.8



**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
State**

Test Date: March 2006
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean					
Total Students²	109,091	689	604	1,351	106,447	25.7	65.7	8.5	214.3	697	685	179	1,07,530	35.5	43.6	20.9	214.6	665	690	136	1,07,600	21.0	58.5	20.5	223.2					
General Education	87,707	0	345	313	87,049	16.0	73.7	10.3	221.4	0	416	66	87,225	26.3	48.7	25.0	222.1	0	414	58	87,235	13.4	62.3	24.2	228.9					
Special Education	18,327	689	216	346	17,076	67.3	32.2	0.5	184.2	697	244	107	17,279	75.0	21.9	3.1	182.9	665	253	76	17,333	50.2	44.9	4.9	200.6					
Limited English Proficient ³	5,188	7	43	701	2,437	84.2	15.6	0.2	170.7	7	26	6	5,149	77.4	18.0	4.7	180.8	6	24	2	5,156	73.7	24.9	1.4	187.2					
Gender⁴																														
Female	52,729	243	255	570	51,661	19.7	68.3	11.9	219.7	249	290	56	52,134	35.9	45.3	18.8	213.7	239	289	45	52,156	21.9	61.0	17.1	221.2					
Male	56,319	444	342	774	54,759	31.3	65.3	5.3	209.2	446	388	120	55,365	35.2	41.9	22.9	215.5	424	395	85	55,415	20.2	56.2	23.7	225.1					
Migrant Status																														
Migrant	38	0	0	2	36	66.7	30.6	2.8	189.1	0	0	0	38	65.8	28.9	5.3	187.8	0	1	0	37	37.8	59.5	2.7	199.6					
Non-Migrant	109,053	689	604	1,349	106,411	25.7	65.8	8.5	214.3	697	685	179	1,07,492	35.5	43.6	20.9	214.6	665	689	136	1,07,563	21.0	58.5	20.5	223.2					
Ethnicity																														
White	62,478	381	194	306	61,597	14.6	74.3	11.1	222.4	382	241	86	61,769	22.1	50.7	27.2	224.9	363	244	54	61,817	9.7	62.7	27.5	232.4					
Black	19,563	125	217	304	18,917	49.9	48.0	2.1	196.4	127	256	52	19,128	68.3	27.4	4.3	188.8	121	257	50	19,135	45.3	50.1	4.6	202.9					
Asian	7,312	36	24	110	7,142	14.5	66.6	18.8	226.3	38	12	4	7,258	14.4	40.8	44.8	237.3	37	13	2	7,260	10.0	53.4	36.7	236.7					
Pacific Islander	305	0	1	0	304	15.1	75.3	9.5	220.2	0	1	0	304	24.3	51.0	24.7	224.0	0	1	1	303	8.6	65.3	26.1	231.3					
Hispanic	18,925	131	156	617	18,021	42.7	54.7	2.6	200.7	134	165	34	18,592	54.5	37.7	7.8	198.2	128	164	23	18,610	38.0	55.1	6.9	208.0					
Amer Indian/AK Native	121	1	0	4	116	38.8	55.2	6.0	205.1	1	0	0	120	46.7	42.5	10.8	203.6	1	0	0	120	28.3	57.5	14.2	216.5					
Other ⁵	387	15	12	10	359	31.4	58.9	9.7	211.0	15	10	3	359	43.5	38.2	18.4	208.5	15	11	6	355	23.7	57.2	19.2	220.7					
Economic Status																														
Econ Disadvantaged	30,006	227	293	746	28,740	49.4	48.8	1.8	196.6	229	332	69	29,376	61.6	32.2	6.3	193.4	217	333	47	29,409	42.9	51.9	5.2	204.5					
Non-Econ Disadvantaged	79,085	462	311	605	77,707	16.9	72.0	11.0	220.9	468	353	110	78,154	25.8	47.9	26.4	222.6	448	357	89	78,191	12.8	61.0	26.2	230.2					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.



**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG A**

Test Date: March 2006
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean					
Total Students²	18,097	95	230	552	17,220	54.3	43.9	1.8	193.4	92	259	38	17,708	66.4	28.0	5.6	189.8	90	264	34	17,709	49.5	46.4	4.1	200.7					
General Education	13,069	0	140	107	12,822	42.7	54.9	2.4	202.4	0	164	11	12,894	58.2	34.6	7.2	196.8	0	163	10	12,896	38.9	55.8	5.3	206.8					
Special Education	3,621	95	77	124	3,525	88.7	11.2	0.1	167.0	92	83	26	3,420	90.4	9.1	0.6	169.0	90	91	24	3,416	77.3	22.1	0.7	184.6					
Limited English Proficient ³	1,487	3	13	328	1,143	86.3	13.6	0.1	167.3	2	13	1	1,471	85.2	14.3	2.5	175.3	2	11	0	1,474	79.6	19.6	0.7	183.3					
Gender⁴																														
Female	8,812	34	84	257	8,437	46.6	50.6	2.8	199.3	34	99	14	8,665	66.0	28.8	5.2	190.0	33	97	10	8,672	50.9	45.6	3.4	199.9					
Male	9,277	61	142	294	8,780	61.7	37.5	0.9	187.8	58	157	24	9,038	66.8	27.3	6.0	189.5	57	165	24	9,031	48.1	47.2	4.7	201.4					
Migrant Status																														
Migrant	31	0	0	2	29	65.5	31.0	3.4	187.0	0	0	0	31	67.7	29.0	3.2	185.2	0	1	0	30	46.7	50.0	3.3	197.2					
Non-Migrant	18,066	95	230	550	17,191	54.3	43.9	1.8	193.4	92	259	38	17,677	66.4	28.0	5.6	189.8	90	263	34	17,679	49.5	46.4	4.1	200.7					
Ethnicity																														
White	1,811	8	11	26	1,766	33.0	63.1	3.9	207.3	8	18	4	1,781	37.8	46.9	15.3	210.5	7	16	2	1,786	23.0	63.0	14.0	218.7					
Black	7,842	41	128	152	7,521	62.1	37.0	0.9	188.6	40	142	23	7,637	78.3	19.5	2.2	181.1	39	150	23	7,630	57.8	40.6	1.7	195.2					
Asian	287	0	1	17	269	38.3	55.8	5.9	206.6	0	0	0	287	39.4	41.8	18.8	210.3	0	0	0	287	32.1	57.5	10.5	213.0					
Pacific Islander	5	0	0	0	5	20.0	80.0	0.0	228.8	0	0	0	5	20.0	60.0	20.0	223.8	0	0	0	5	20.0	80.0	0.0	223.0					
Hispanic	8,091	42	85	354	7,610	52.0	45.9	2.1	194.6	40	96	11	7,944	62.4	31.5	6.2	192.7	40	95	9	7,947	48.1	47.9	3.9	201.4					
Amer. Indian/AK Native	18	0	0	1	17	47.1	41.2	11.8	202.5	0	0	0	18	38.9	50.0	11.1	207.9	0	0	0	18	44.4	50.0	5.6	212.1					
Other ⁵	43	4	5	2	32	81.3	18.8	0.0	175.1	4	3	0	36	86.1	13.9	0.0	174.0	4	3	0	36	63.9	36.1	0.0	188.7					
Economic Status																														
Econ Disadvantaged	15,158	60	159	385	12,554	57.2	41.4	1.4	191.7	58	179	22	12,899	69.1	26.2	4.7	187.9	56	180	19	12,903	52.5	44.5	3.0	198.7					
Non-Econ Disadvantaged	4,939	35	71	167	4,666	46.5	50.8	2.7	198.2	34	80	16	4,809	59.0	33.0	7.9	194.7	34	84	15	4,806	41.5	51.4	7.1	205.9					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.



**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG B**

Test Date: March 2006
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voils ¹	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voils	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voils	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean					
Total Students²	11,440	107	74	231	11,028	37.6	59.9	3.5	204.5	111	94	27	11,208	48.1	40.1	11.9	203.8	107	96	25	11,212	30.4	59.2	10.4	213.3					
General Education	8,553	0	32	38	8,483	24.4	71.0	4.6	213.4	0	46	2	8,505	37.1	47.9	15.0	212.3	0	41	7	8,505	19.8	67.0	13.1	219.9					
Special Education	2,312	107	40	55	2,110	80.6	19.4	0.0	175.6	111	40	24	2,137	83.7	14.6	1.6	176.5	107	49	17	2,139	61.6	36.5	1.9	193.8					
Limited English Proficient ³	588	1	2	138	447	86.8	13.2	0.0	170.9	2	8	1	577	78.9	18.4	2.8	179.0	1	6	1	580	72.2	27.1	0.7	187.6					
Gender⁴																														
Female	5,560	42	24	92	5,382	30.4	64.5	5.1	209.8	45	38	3	5,454	49.7	40.2	10.0	202.6	42	35	6	5,457	32.6	59.2	8.2	211.0					
Male	5,895	64	50	138	5,643	44.5	53.5	2.0	199.4	65	56	23	5,751	46.5	39.9	13.6	204.9	64	61	18	5,752	28.4	59.2	12.4	215.4					
Migrant Status																														
Migrant	1	0	0	0	1	100.0	0.0	0.0	183.0	0	0	0	1	0.0	0.0	100.0	255.0	0	0	0	1	0.0	100.0	0.0	218.0					
Non-Migrant	11,439	107	74	231	11,027	37.6	58.9	3.5	204.5	111	94	27	11,207	48.1	40.1	11.9	203.8	107	96	25	11,211	30.5	59.2	10.4	213.3					
Ethnicity																														
White	4,836	52	27	49	4,708	27.0	67.5	5.5	212.0	52	34	7	4,743	35.0	47.4	17.6	213.6	51	37	8	4,740	18.3	65.8	16.0	222.2					
Black	2,744	22	32	54	2,636	52.7	46.2	1.1	194.1	23	40	13	2,668	70.9	26.0	3.1	187.6	22	39	11	2,672	47.7	49.3	3.0	200.9					
Asian	574	3	1	23	547	28.7	63.8	7.5	212.1	4	1	0	569	27.1	49.2	23.7	220.3	3	1	0	570	17.7	63.9	18.4	223.4					
Pacific Islander	46	0	1	0	45	22.2	64.4	13.3	215.6	0	1	0	45	48.9	26.7	24.4	212.4	0	1	0	45	20.0	64.4	15.6	222.4					
Hispanic	3,169	28	13	103	3,025	42.7	55.7	1.7	200.3	30	18	7	3,114	52.4	39.3	8.3	199.7	29	18	6	3,116	36.6	56.8	6.5	208.3					
Amer Indian/AK Native	35	0	0	2	33	42.4	54.5	3.0	204.7	0	0	0	35	40.0	57.1	2.9	199.7	0	0	0	35	34.3	60.0	5.7	210.9					
Other ⁵	36	2	0	0	34	41.2	50.0	8.8	205.3	2	0	0	34	44.1	38.2	17.6	207.4	2	0	0	34	29.4	47.1	23.5	217.9					
Economic Status																														
Econ Disadvantaged	5,849	60	42	165	5,582	47.6	50.8	1.6	197.4	62	52	15	5,720	58.3	34.5	7.2	195.7	60	57	12	5,720	40.3	54.3	5.5	206.1					
Non-Econ Disadvantaged	5,591	47	32	66	5,446	27.4	67.1	5.5	211.8	49	42	12	5,488	37.4	45.8	16.7	212.2	47	39	13	5,492	20.2	64.3	15.5	220.8					

¹ Include students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.



**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG CD**

Test Date: March 2006
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean					
Total Students²	10,556	84	75	137	10,260	29.4	64.7	5.9	210.7	82	84	27	10,363	40.0	45.3	14.7	209.6	83	81	11	10,381	21.6	64.0	14.4	220.2					
General Education	8,389	0	39	38	8,312	18.1	74.7	7.2	218.1	0	45	12	8,332	29.9	52.4	17.8	217.3	0	44	3	8,342	13.0	69.7	17.3	226.0					
Special Education	1,938	84	54	44	1,776	76.6	25.1	0.2	179.9	82	38	13	1,805	81.9	16.2	1.9	177.6	83	37	7	1,811	54.5	43.0	2.5	197.6					
Limited English Proficient ³	237	1	2	55	179	86.6	13.4	0.0	169.3	1	1	2	233	79.8	16.7	3.4	179.8	1	0	1	235	76.2	22.1	1.7	186.6					
Gender⁴																														
Female	5,131	33	34	45	5,019	21.0	70.5	8.5	216.8	32	39	6	5,054	39.5	47.2	13.3	209.5	32	35	3	5,061	21.8	66.3	12.0	218.8					
Male	5,422	51	41	92	5,238	37.5	59.2	3.3	204.9	50	45	20	5,307	40.5	43.4	16.1	209.6	51	46	8	5,317	21.4	61.9	16.7	221.5					
Migrant Status																														
Migrant	1	0	0	0	1	0.0	100.0	0.0	200.0	0	0	0	1	0.0	100.0	0.0	218.0	0	0	0	1	0.0	100.0	0.0	208.0					
Non-Migrant	10,555	84	75	137	10,259	29.4	64.7	5.9	210.7	82	84	27	10,362	40.0	45.3	14.7	209.6	83	81	11	10,380	21.6	64.0	14.4	220.2					
Ethnicity																														
White	5,862	40	28	45	5,771	21.8	70.4	7.8	216.4	38	35	18	5,791	29.5	50.4	20.1	217.7	39	34	3	5,806	13.4	66.6	20.0	227.1					
Black	2,051	25	25	30	1,971	43.9	53.5	2.6	200.6	25	28	5	1,993	62.5	33.0	4.6	195.2	25	25	5	1,996	36.8	58.2	5.0	207.8					
Asian	497	2	3	10	482	23.4	65.4	11.2	216.7	2	3	2	490	21.8	52.2	25.9	223.9	2	3	0	492	15.0	65.0	19.9	226.3					
Pacific Islander	86	0	0	0	86	14.0	77.9	8.1	219.1	0	0	0	86	22.1	57.0	20.9	222.5	0	0	0	86	8.1	70.9	20.9	227.7					
Hispanic	1,995	17	18	54	1,906	39.1	59.0	1.9	202.5	17	18	1	1,959	53.0	40.8	6.2	198.6	17	19	3	1,956	32.4	62.0	5.6	210.7					
Amer Indian/AK Native	12	0	0	0	12	50.0	50.0	0.0	196.5	0	0	0	12	75.0	16.7	8.3	186.2	0	0	0	12	50.0	41.7	8.3	205.2					
Other ⁵	33	0	1	0	32	56.3	37.5	6.3	189.5	0	0	1	32	65.6	34.4	0.0	185.0	0	0	0	33	39.4	51.5	9.1	204.5					
Economic Status																														
Econ Disadvantaged	3,262	32	34	67	3,129	44.5	53.4	2.1	200.1	31	40	11	3,180	55.8	37.6	6.6	197.1	32	36	4	3,190	34.1	59.4	6.5	209.5					
Non-Econ Disadvantaged	7,294	52	41	70	7,131	22.8	69.7	7.5	215.3	51	44	16	7,183	33.0	48.7	18.3	215.1	51	45	7	7,191	16.1	66.0	17.9	224.9					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.

Test Date: March 2006
CYCLE II

New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG DE



	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voits ¹	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voits	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voits	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean					
Total Students²	14,160	124	68	107	13,861	22.7	70.8	6.5	215.0	122	72	28	13,938	33.5	48.1	18.4	215.1	117	76	12	13,955	16.7	63.5	19.9	225.1					
General Education	11,527	0	39	50	11,438	13.2	78.9	7.8	221.3	0	44	12	11,471	24.4	53.7	21.9	221.7	0	50	3	11,474	10.3	66.6	23.2	229.9					
Special Education	2,436	124	22	34	2,256	65.9	33.9	0.2	186.4	122	28	15	2,271	75.4	22.4	2.2	184.5	117	24	9	2,286	44.0	51.3	4.7	203.7					
Limited English Proficient ³	203	0	7	24	172	87.2	12.2	0.6	169.4	0	0	1	202	77.7	18.8	3.5	181.7	0	2	0	201	72.6	25.9	1.5	188.9					
Gender⁴																														
Female	6,876	38	33	39	6,766	15.8	74.8	9.4	220.6	37	30	8	6,801	33.8	49.9	16.2	214.2	37	34	3	6,802	17.4	66.0	16.6	223.0					
Male	7,276	86	35	67	7,090	29.2	67.0	3.8	209.7	85	41	20	7,130	33.1	46.4	20.5	216.0	80	40	9	7,147	15.9	61.1	23.0	227.0					
Migrant Status																														
Migrant	0	0	0	0	0					0	0	0	0						0	0	0	0								
Non-Migrant	14,160	124	68	107	13,861	22.7	70.8	6.5	215.0	122	72	28	13,938	33.5	48.1	18.4	215.1	117	76	12	13,955	16.7	63.5	19.9	225.1					
Ethnicity																														
White	9,809	92	48	56	9,613	17.6	74.9	7.5	218.6	89	50	21	9,649	26.5	52.3	21.3	220.1	87	49	8	9,665	11.7	65.0	23.3	229.5					
Black	1,899	14	6	21	1,858	39.5	57.9	2.6	202.7	14	13	2	1,870	62.4	32.1	5.5	193.8	13	13	2	1,871	33.9	59.8	6.3	209.1					
Asian	743	6	4	5	728	21.7	67.9	10.4	219.0	6	1	0	736	18.5	46.6	34.9	229.7	6	2	0	735	15.1	57.4	27.5	230.2					
Pacific Islander	28	0	0	0	28	21.4	67.9	10.7	215.2	0	0	0	28	28.6	53.6	17.9	216.9	0	0	0	28	10.7	60.7	28.6	227.7					
Hispanic	1,609	11	7	24	1,567	34.5	62.8	2.7	205.4	12	4	5	1,588	48.8	42.6	8.6	202.7	10	6	2	1,591	27.8	61.3	10.9	214.5					
Amer Indian/AK Native	7	0	0	0	7	0.0	100.0	0.0	211.7	0	0	0	7	42.9	42.9	14.3	207.3	0	0	0	7	0.0	71.4	28.6	229.0					
Other ⁵	65	1	3	1	60	23.3	63.3	13.3	220.1	1	4	0	60	36.7	40.0	23.3	215.7	1	6	0	58	8.6	72.4	19.0	228.2					
Economic Status																														
Econ Disadvantaged	2,660	50	22	34	2,574	40.2	57.8	2.0	202.3	30	22	8	2,600	54.3	37.5	8.3	199.1	29	22	3	2,606	30.8	60.3	8.9	212.0					
Non-Econ Disadvantaged	11,500	94	46	73	11,287	18.7	73.8	7.5	217.9	92	50	20	11,338	28.7	50.6	20.8	218.8	88	54	9	11,349	13.4	64.2	22.4	228.1					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.



**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG FG**

Test Date: March 2006
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean					
Total Students²	14,649	89	37	97	14,426	18.4	73.9	7.7	218.4	89	54	10	14,496	28.0	51.3	20.8	218.9	81	51	13	14,504	12.8	64.0	23.1	228.8					
General Education	12,199	0	24	28	12,147	9.9	81.0	9.1	224.0	0	34	5	12,160	19.5	56.3	24.2	225.3	0	33	8	12,158	7.0	66.3	26.7	233.5					
Special Education	2,282	89	9	28	2,156	62.8	36.9	0.3	188.8	89	17	5	2,171	71.9	25.4	2.6	185.8	81	15	5	2,181	41.1	53.9	5.0	205.2					
Limited English Proficient ³	174	0	4	41	129	79.8	19.4	0.8	177.1	0	3	0	171	73.7	21.6	4.7	185.1	0	3	0	171	69.6	28.7	1.8	190.8					
Gender⁴																														
Female	7,049	21	18	43	6,967	12.2	76.5	11.3	224.2	22	22	2	7,003	27.9	53.4	18.7	218.0	20	24	4	7,001	13.1	67.6	19.3	226.7					
Male	7,596	67	19	53	7,457	24.2	71.4	4.4	213.0	66	31	8	7,491	28.0	49.3	22.7	219.8	60	26	9	7,501	12.5	60.7	26.8	230.7					
Migrant Status																														
Migrant	1	0	0	0	1	100.0	0.0	0.0	181.0	0	0	0	1	100.0	0.0	0.0	162.0	0	0	0	1	0.0	100.0	0.0	202.0					
Non-Migrant	14,648	89	37	97	14,425	18.4	73.9	7.8	218.4	89	54	10	14,495	28.0	51.3	20.8	218.9	81	51	13	14,503	12.8	64.0	23.1	228.8					
Ethnicity																														
White	11,061	61	25	51	10,924	15.6	76.1	8.3	220.4	59	37	6	10,959	24.2	53.2	22.6	221.7	54	38	10	10,959	9.7	64.9	25.4	231.5					
Black	1,176	7	2	14	1,153	33.5	63.5	3.0	206.7	8	5	1	1,162	55.2	38.2	6.5	197.9	7	4	2	1,163	29.6	61.4	9.0	212.9					
Asian	869	7	3	9	850	15.2	71.1	13.8	222.9	7	1	0	861	15.7	49.0	35.3	232.3	7	1	0	861	10.2	59.9	29.8	233.5					
Pacific Islander	106	0	0	0	106	13.2	77.4	9.4	222.7	0	0	0	106	17.9	57.5	24.5	228.5	0	0	1	105	3.8	61.0	35.2	238.2					
Hispanic	1,380	11	7	23	1,339	30.2	66.4	3.4	209.3	12	11	3	1,354	43.1	48.4	8.6	205.2	10	8	0	1,362	25.9	62.6	11.5	216.3					
Amer Indian/AK Native	17	1	0	0	16	43.8	50.0	6.3	206.4	1	0	0	16	50.0	31.3	18.8	207.7	1	0	0	16	25.0	50.0	25.0	221.8					
Other ⁵	40	2	0	0	38	18.4	76.3	5.3	218.8	2	0	0	38	26.3	44.7	28.9	221.9	2	0	0	38	15.8	50.0	34.2	236.2					
Economic Status																														
Econ Disadvantaged	1,831	18	9	27	1,777	35.6	61.7	2.8	205.7	19	14	2	1,796	47.8	43.6	8.6	202.5	15	11	1	1,804	29.2	62.7	8.1	213.2					
Non-Econ Disadvantaged	12,818	71	28	70	12,649	16.0	75.6	8.5	220.2	70	40	8	12,700	25.2	52.3	22.5	221.2	66	40	12	12,700	10.5	64.2	25.3	231.0					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.

**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG GH**



	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean					
Total Students²	14,124	80	56	90	13,898	14.1	73.7	12.2	223.1	87	56	7	13,974	22.7	48.2	29.1	225.5	74	55	10	13,985	11.3	62.1	26.7	231.4					
General Education	11,840	0	35	17	11,788	6.9	78.9	14.2	228.4	0	40	4	11,796	15.2	51.4	33.4	231.6	0	36	7	11,797	6.2	63.4	30.4	235.8					
Special Education	2,072	80	15	21	1,956	52.9	46.1	1.0	194.6	87	15	3	1,967	63.4	31.0	5.6	192.2	74	18	3	1,977	36.5	56.4	7.0	208.7					
Limited English Proficient ³	218	2	6	53	157	76.4	22.3	1.3	178.9	2	1	0	215	60.0	29.3	10.7	194.5	2	1	0	215	59.5	37.7	2.8	195.7					
Gender⁴																														
Female	6,777	27	30	34	6,686	9.4	73.9	16.8	228.3	28	29	4	6,716	23.0	50.8	26.2	224.1	26	25	6	6,720	12.1	65.8	22.1	228.8					
Male	7,340	53	26	55	7,206	18.5	73.5	8.0	218.3	59	27	3	7,251	22.3	45.8	31.9	226.9	48	30	4	7,258	10.4	58.6	31.0	233.8					
Migrant Status																														
Migrant	1	0	0	0	1	100.0	0.0	0.0	192.0	0	0	0	1	100.0	0.0	0.0	189.0	0	0	0	1	0.0	100.0	0.0	200.0					
Non-Migrant	14,123	80	56	90	13,897	14.1	73.7	12.2	223.1	87	56	7	13,973	22.7	48.2	29.1	225.5	74	55	10	13,984	11.3	62.0	26.7	231.4					
Ethnicity																														
White	9,820	50	26	33	9,711	10.4	76.8	12.8	225.7	56	30	5	9,729	17.2	51.4	31.4	229.3	47	30	7	9,736	7.2	63.6	29.2	234.7					
Black	1,531	8	12	9	1,502	30.3	64.2	5.5	209.7	9	13	1	1,508	50.9	40.8	8.3	201.1	7	15	2	1,507	28.5	61.6	9.9	214.2					
Asian	1,524	9	3	22	1,490	11.9	67.0	21.1	228.9	9	2	0	1,513	12.2	39.5	48.4	240.3	9	1	1	1,513	8.7	52.1	39.1	238.7					
Pacific Islander	7	0	0	0	7	0.0	85.7	14.3	225.1	0	0	0	7	42.9	42.9	14.3	215.3	0	0	0	7	14.3	57.1	28.6	219.3					
Hispanic	1,170	12	13	23	1,122	27.1	68.0	4.9	211.1	12	9	1	1,148	44.4	43.7	11.8	206.5	10	8	0	1,152	25.7	62.7	11.6	216.6					
Amer Indian/AK Native	16	0	0	0	16	37.5	56.3	6.3	204.3	0	0	0	16	56.3	31.3	12.5	200.4	0	0	0	16	12.5	68.8	18.8	221.1					
Other ⁵	56	1	2	3	50	18.0	72.0	10.0	221.3	1	2	0	53	39.6	30.2	30.2	216.2	1	1	0	54	20.4	57.4	22.2	226.7					
Economic Status																														
Econ Disadvantaged	1,353	18	14	25	1,296	36.6	60.5	2.9	205.8	20	10	1	1,322	50.9	38.8	10.3	202.6	15	14	0	1,324	30.5	60.2	9.3	212.7					
Non-Econ Disadvantaged	12,771	62	42	65	12,602	11.8	75.0	13.2	224.9	67	46	6	12,652	19.7	49.2	31.1	227.9	59	41	10	12,661	9.3	62.2	28.5	233.3					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.



**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG I**

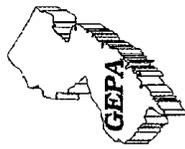
Test Date: March 2006
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voils ¹	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voils	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voils	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean					
Total Students²	19,987	89	42	78	19,778	9.8	74.5	15.8	227.9	93	41	20	19,833	17.3	47.7	35.1	231.2	92	42	12	19,841	7.6	58.2	34.2	236.7					
General Education	17,064	0	23	21	17,020	4.2	77.8	18.0	232.5	0	29	9	17,026	10.7	49.7	39.6	236.8	0	31	7	17,026	3.7	58.1	38.2	240.6					
Special Education	2,756	89	11	14	2,622	43.1	55.1	1.7	200.0	93	12	10	2,621	57.4	36.0	6.6	196.4	92	10	5	2,629	29.1	60.1	10.8	214.1					
Limited English Proficient ³	195	0	8	43	144	66.0	33.3	0.7	185.8	0	0	1	194	53.1	28.4	18.6	202.7	0	1	0	194	55.7	39.2	5.2	199.6					
Gender⁴																														
Female	9,598	38	23	43	9,494	6.0	72.5	21.5	233.1	41	23	11	9,523	17.8	50.2	32.0	229.6	39	26	9	9,524	7.9	63.2	28.9	234.1					
Male	10,388	51	18	35	10,284	13.2	76.3	10.4	223.1	52	17	9	10,310	16.8	45.3	37.9	232.6	53	15	3	10,317	7.3	53.6	39.1	239.1					
Migrant Status																														
Migrant	0	0	0	0	0					0	0	0	0						0	0	0	0								
Non-Migrant	19,987	89	42	78	19,778	9.8	74.5	15.8	227.9	93	41	20	19,833	17.3	47.7	35.1	231.2	92	42	12	19,841	7.6	58.2	34.2	236.7					
Ethnicity																														
White	15,745	67	22	37	15,619	7.9	76.5	15.7	229.0	69	30	16	15,630	15.0	49.9	35.1	232.3	67	33	11	15,634	5.9	59.0	35.1	238.1					
Black	1,136	6	6	3	1,121	30.1	64.9	5.0	210.5	6	4	0	1,126	48.5	40.1	11.5	203.6	6	3	0	1,127	26.3	61.2	12.5	216.3					
Asian	2,118	8	7	19	2,084	7.1	67.4	25.6	234.2	9	2	2	2,105	8.1	35.4	56.5	246.6	9	3	1	2,105	4.7	47.9	47.4	244.3					
Pacific Islander	11	0	0	0	11	9.1	72.7	18.2	227.0	0	0	0	11	0.0	36.4	63.6	247.1	0	0	0	11	0.0	63.6	36.4	247.4					
Hispanic	889	7	6	16	860	23.3	69.1	7.7	215.0	8	4	2	875	38.4	47.0	14.6	210.2	9	2	0	878	20.4	63.9	15.7	221.3					
Amer Indian/AK Native	13	0	0	1	12	33.3	58.3	8.3	207.4	0	0	0	13	38.5	46.2	15.4	213.9	0	0	0	13	15.4	61.5	23.1	221.1					
Other ⁵	75	1	1	2	71	14.1	67.6	18.3	224.7	1	1	0	73	26.0	56.2	17.8	219.3	1	1	0	73	11.0	67.1	21.9	228.7					
Economic Status																														
Econ Disadvantaged	781	4	3	9	765	35.4	61.4	3.1	206.1	4	3	1	773	48.9	42.4	8.7	202.3	5	1	0	775	28.5	62.5	9.0	214.0					
Non-Econ Disadvantaged	19,206	85	39	69	19,013	8.7	75.0	16.3	228.7	89	38	19	19,060	16.0	47.9	36.2	232.3	87	41	12	19,066	6.7	58.0	35.3	237.6					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.

Test Date: March 2006
CYCLE II

**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG J**



	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	Advanced Proficient	Scale Score Mean					
Total Students²	4,190	15	7	19	4,149	6.3	71.8	21.9	232.9	15	6	9	4,160	10.5	44.1	45.4	239.7	15	4	8	4,163	4.0	51.8	44.2	243.0					
General Education	3,574	0	5	2	3,567	1.8	73.0	25.2	237.3	0	5	4	3,565	4.5	44.4	51.2	245.3	0	2	5	3,567	1.2	50.0	48.8	246.7					
Special Education	572	15	1	7	549	32.2	66.1	1.6	207.2	15	1	5	551	46.5	43.9	9.6	205.1	15	2	3	552	19.4	63.6	17.0	221.5					
Limited English Proficient ³	47	0	1	10	36	69.4	30.6	0.0	183.6	0	0	0	47	48.9	25.5	25.5	213.8	0	0	0	47	42.6	51.1	6.4	206.3					
Gender⁴																														
Female	2,052	6	4	8	2,034	4.1	66.1	29.8	237.9	6	4	4	2,038	11.0	46.7	42.3	237.9	6	2	2	2,042	4.5	56.9	36.6	240.3					
Male	2,134	9	3	11	2,111	8.4	77.3	14.3	228.2	9	2	5	2,118	10.0	41.6	48.4	241.3	9	2	2	2,121	3.5	47.0	49.5	245.5					
Migrant Status																														
Migrant	0	0	0	0	0					0	0	0	0					0	0	0	0	0								
Non-Migrant	4,190	15	7	19	4,149	6.3	71.8	21.9	232.9	15	6	9	4,160	10.5	44.1	45.4	239.7	15	4	8	4,163	4.0	51.8	44.2	243.0					
Ethnicity																														
White	3,277	11	5	9	3,252	5.3	73.3	21.4	233.4	11	5	7	3,254	9.8	47.3	42.9	238.7	11	3	3	3,260	3.4	52.9	43.7	243.1					
Black	81	0	0	0	81	16.0	77.8	6.2	221.1	0	0	0	81	40.7	44.4	14.8	210.6	0	0	0	81	18.5	63.0	18.5	224.6					
Asian	659	1	1	5	652	7.4	63.2	29.4	235.3	1	1	0	657	6.1	27.4	66.5	252.5	1	1	0	657	3.3	41.6	55.1	248.2					
Pacific Islander	16	0	0	0	16	12.5	87.5	0.0	222.6	0	0	0	16	12.5	50.0	37.5	235.1	0	0	0	16	6.3	75.0	18.8	233.4					
Hispanic	142	1	1	5	135	18.5	71.9	9.6	219.7	1	0	2	139	27.3	48.9	25.7	219.9	1	0	1	140	12.9	65.0	22.1	227.3					
Amer Indian/AK Native	3	0	0	0	3	0.0	66.7	33.3	232.0	0	0	0	3	33.3	33.3	33.3	234.3	0	0	0	3	0.0	66.7	33.3	250.0					
Other ⁵	12	2	0	0	10	30.0	60.0	10.0	215.9	2	0	0	10	10.0	40.0	50.0	244.9	2	0	4	6	0.0	50.0	50.0	250.5					
Economic Status																														
Econ Disadvantaged	60	0	0	2	58	27.6	65.5	6.9	213.0	0	0	1	59	42.4	45.8	11.9	211.4	0	0	1	59	22.0	64.4	13.6	219.1					
Non-Econ Disadvantaged	4,130	15	7	17	4,091	6.0	71.8	22.1	233.2	15	6	8	4,101	10.0	44.1	45.9	240.1	15	4	7	4,104	3.8	51.6	44.6	243.3					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.



**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
DFG R**

Test Date: March 2006
CYCLE II

	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean					
Total Students²	1,358	5	3	8	1,342	43.1	54.0	2.9	201.1	5	5	5	1,343	60.6	31.3	8.1	195.9	5	5	2	1,346	37.3	52.2	10.5	210.2					
General Education	1,163	0	2	3	1,158	36.2	60.4	3.4	206.1	0	3	2	1,158	55.6	35.0	9.4	199.8	0	3	1	1,159	31.8	56.3	11.9	213.7					
Special Education	190	5	1	5	179	86.0	14.0	0.0	169.7	5	2	3	180	91.7	8.3	0.0	171.4	5	2	1	182	72.5	25.3	2.2	187.8					
Limited English Proficient ³	6	0	0	0	6	100.0	0.0	0.0	155.8	0	0	0	6	100.0	0.0	0.0	173.5	0	0	0	6	50.0	50.0	0.0	191.0					
Gender⁴																														
Female	698	3	2	4	689	40.3	55.4	4.2	204.1	3	3	3	689	61.4	32.5	6.1	194.3	3	4	1	690	39.9	52.9	7.2	207.1					
Male	658	2	1	2	653	43.9	52.5	1.5	198.0	2	1	1	654	59.8	30.0	10.2	197.6	2	1	0	655	34.5	51.5	14.0	213.4					
Migrant Status																														
Migrant	3	0	0	0	3	66.7	35.3	0.0	209.0	0	0	0	3	66.7	33.3	0.0	190.0	0	0	0	3	0.0	100.0	0.0	213.7					
Non-Migrant	1,355	5	3	8	1,339	43.0	54.1	2.9	201.1	5	5	5	1,340	60.6	31.3	8.1	195.9	5	5	2	1,343	37.4	52.0	10.6	210.2					
Ethnicity																														
White	152	0	2	1	149	22.1	68.5	9.4	217.3	0	2	1	149	31.5	38.3	30.2	223.0	0	2	1	149	13.4	49.0	37.6	234.7					
Black	864	1	0	6	857	51.0	47.3	1.8	195.8	1	2	4	857	70.9	26.1	2.9	187.6	1	2	1	860	46.0	49.0	5.0	203.0					
Asian	36	0	0	0	36	8.3	86.1	5.6	220.3	0	0	0	36	13.9	50.0	36.1	231.8	0	0	0	36	11.1	44.4	44.4	237.6					
Pacific Islander	0	0	0	0	0					0	0	0	0					0	0	0	0									
Hispanic	287	2	1	1	283	35.7	61.5	2.8	205.8	2	1	0	284	50.7	40.5	8.8	202.3	2	1	0	284	27.5	63.7	8.8	215.2					
Amer Indian/AK Native	0	0	0	0	0					0	0	0	0					0	0	0	0									
Other ⁵	19	2	0	0	17	23.5	76.5	0.0	208.1	2	0	0	17	58.8	35.3	5.9	195.3	2	0	0	17	23.5	64.7	11.8	214.4					
Economic Status																														
Econ Disadvantaged	731	5	1	6	719	48.0	50.3	1.7	197.1	5	3	4	719	65.9	30.2	3.9	190.8	5	3	1	722	43.4	50.7	6.0	205.0					
Non-Econ Disadvantaged	627	0	2	2	623	37.4	58.3	4.3	205.7	0	2	1	624	54.5	32.5	13.0	201.8	0	2	1	624	30.3	53.8	15.9	216.2					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.

**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
Special Needs**

Test Date: March 2006
CYCLE II



	Language Arts Literacy										Mathematics										Science									
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean					
Total Students²	21,372	127	262	642	20,341	52.7	45.5	1.8	194.4	128	299	47	20,898	65.1	28.8	6.1	190.8	123	296	44	20,909	48.2	47.4	4.5	201.6					
General Education	15,340	0	152	118	15,070	40.4	57.2	2.4	203.7	0	184	9	15,147	56.4	35.7	7.9	198.3	0	180	11	15,149	37.1	57.0	5.9	207.9					
Special Education	4,414	127	95	140	4,052	88.4	11.5	0.1	167.7	128	99	37	4,150	90.1	9.3	0.7	169.3	123	104	32	4,155	76.7	22.7	0.6	185.2					
Limited English Proficient ³	1,699	3	15	390	1,291	86.4	13.5	0.1	168.0	3	17	1	1,678	82.6	15.1	2.3	175.8	2	13	1	1,683	78.8	20.5	0.7	183.8					
Gender⁴																														
Female	10,454	48	96	286	10,024	45.0	52.3	2.7	200.2	51	117	14	10,272	65.1	29.4	5.5	190.9	47	107	12	10,288	49.7	46.7	3.6	200.6					
Male	10,908	78	162	355	10,313	60.1	38.9	1.0	188.8	76	179	33	10,620	65.0	28.3	6.7	190.7	75	187	32	10,614	46.7	48.0	5.4	202.5					
Migrant Status																														
Migrant	28	0	0	0	28	64.3	32.1	3.6	187.9	0	0	0	28	67.9	28.6	3.6	186.5	0	1	0	27	44.4	51.9	3.7	197.7					
Non-Migrant	21,344	127	262	642	20,313	52.7	45.5	1.8	194.4	128	299	47	20,870	65.1	28.8	6.1	190.8	123	295	44	20,882	48.2	47.4	4.5	201.6					
Ethnicity																														
White	2,402	13	15	43	2,331	34.1	62.2	3.6	206.7	13	25	5	2,359	40.7	44.0	15.3	208.3	12	20	6	2,364	24.4	62.3	13.3	217.9					
Black	9,181	55	148	178	8,800	60.4	38.6	0.9	189.5	55	164	29	8,933	77.1	20.5	2.4	182.1	53	171	29	8,928	56.7	41.4	1.8	195.9					
Asian	555	2	2	27	524	32.4	60.9	6.7	208.9	3	1	0	551	32.7	45.7	21.6	215.8	2	1	0	552	24.1	61.6	14.3	218.5					
Pacific Islander	42	0	1	0	41	26.8	58.5	14.6	217.4	0	1	0	41	46.3	29.3	24.4	212.8	0	1	0	41	17.1	65.9	17.1	223.2					
Hispanic	9,096	56	91	389	8,560	51.0	47.1	1.9	195.2	56	105	13	8,922	61.6	32.1	6.3	193.2	55	100	9	8,932	47.5	48.4	4.1	201.8					
Amer Indian/AK Native	40	0	0	3	37	48.6	45.9	5.4	202.5	0	0	0	40	42.5	52.5	5.0	200.1	0	0	0	40	42.5	52.5	5.0	208.3					
Other ⁵	56	1	5	2	48	70.8	25.0	4.2	183.6	1	3	0	52	76.9	21.2	1.9	180.7	1	3	0	52	53.8	40.4	5.8	196.6					
Economic Status																														
Econ Disadvantaged	15,581	84	185	459	14,833	55.4	43.1	1.5	192.6	84	210	28	15,259	67.7	27.2	5.1	189.0	81	209	28	15,263	51.1	45.6	3.3	199.6					
Non-Econ Disadvantaged	5,791	43	77	183	5,488	45.2	51.9	2.8	193.3	44	89	19	5,639	58.0	33.3	8.7	195.7	42	87	16	5,646	40.1	52.2	7.7	206.9					

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.



**New Jersey Statewide Testing System
Grade Eight Proficiency Assessment
Performance by Demographic Groups
Non Special Needs**

Test Date: March 2006
CYCLE II

	Language Arts Literacy						Mathematics						Science					
	Students Enrolled	APA Students	Not Present	Voids ¹	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	APA Students	Not Present	Voids	Valid Scale Scores	% Partially Proficient	% Proficient	% Advanced Proficient	Scale Score Mean	
Total Students²	87,719	562	342	709	86,106	19.3	70.5	10.1	219.0	569	386	132	86,632	28.4	47.1	24.5	220.4	
General Education	72,367	0	193	195	71,979	10.9	77.1	12.0	225.1	0	232	57	72,078	20.0	51.5	28.6	227.1	
Special Education	13,913	562	121	206	13,024	60.7	38.6	0.7	189.3	569	145	70	13,129	70.2	26.0	3.9	187.2	
Limited English Proficient ³	1,489	4	28	311	1,146	81.7	17.9	0.4	173.7	4	9	5	1,471	71.4	21.2	7.4	186.6	
Gender⁴																		
Female	42,275	195	159	284	41,637	13.7	72.2	14.1	224.5	198	173	42	41,862	28.7	49.2	22.1	219.3	
Male	45,411	366	180	419	44,446	24.6	69.0	6.4	213.9	370	209	87	44,745	28.2	45.1	26.7	221.4	
Migrant Status																		
Migrant	10	0	0	2	8	75.0	25.0	0.0	193.1	0	0	0	10	60.0	30.0	10.0	191.2	
Non-Migrant	87,709	562	342	707	86,098	19.3	70.5	10.1	219.0	569	386	132	86,622	28.4	47.1	24.5	220.4	
Ethnicity																		
White	60,076	368	179	263	59,266	13.8	74.8	11.3	223.0	369	216	81	59,410	21.4	50.9	27.7	225.5	
Black	10,382	70	69	126	10,117	40.8	56.2	3.0	202.4	72	92	23	10,195	60.6	33.5	5.9	194.6	
Asian	6,757	34	22	83	6,618	13.1	67.1	19.8	227.7	35	11	4	6,707	12.9	40.4	46.7	239.1	
Pacific Islander	265	0	0	0	263	13.3	77.9	8.7	220.7	0	0	0	263	20.9	54.4	24.7	225.7	
Hispanic	9,829	75	65	228	9,461	35.1	61.5	3.3	205.6	78	60	21	9,670	48.0	42.8	9.2	202.8	
Amer Indian/AK Native	81	1	0	1	79	34.2	59.5	6.3	206.4	1	0	0	80	48.8	37.5	13.8	205.4	
Other ⁵	331	14	7	8	302	25.2	64.2	10.6	215.4	14	7	3	307	37.8	41.0	21.2	213.2	
Economic Status																		
Econ Disadvantaged	14,425	143	108	287	13,887	43.0	54.8	2.2	200.8	145	122	41	14,117	55.0	37.5	7.5	198.2	
Non-Econ Disadvantaged	73,294	419	234	422	72,219	14.8	73.6	11.7	222.5	424	264	91	72,515	23.2	49.0	27.8	224.7	

¹ Includes students coded as LEP exempt (LAL only).
² Students appear in each applicable category, but they are included in Total Students only once.
³ Excludes students coded as Former LEP.
⁴ Excludes students who did not have Gender coded.
⁵ Includes students who did not have an ethnicity coded and students who had more than one ethnicity coded.

Appendix C

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

In accordance with the Individuals with Disabilities Education Act (IDEA), students who are receiving special education services must participate in each subject area of the age-appropriate statewide assessment with the following exception:

Students with disabilities shall participate in the Alternate Proficiency Assessment in each content area where the nature of the student's disability is so severe that the student is not receiving instruction in any of the knowledge and skills measured by the general statewide assessment and the student cannot complete any of the types of questions on the assessment content area(s) even with accommodation and modifications. (New Jersey Administrative Code Chapter 6A:14-4.11[a]2)

Districts may use modifications of test administration procedures when administering the GEPA to special education students or to students eligible under Section 504 of the Rehabilitation Act of 1973. Decisions about participation and accommodations/modifications are made by the Individualized Education Program (IEP) or 504 team. Information about test content and item types from the test specifications booklets can be used to make this determination. Modifications in the areas listed below may be used separately or in combination.

Any accommodations or modifications of test administration procedures for students eligible for special education under the IDEA or eligible under Section 504 of the Rehabilitation Act of 1973 must be specified in the student's IEP or 504 accommodation plan. Accommodations or modifications must be consistent with the instruction and assessment procedures used in the student's classroom. Students eligible for modifications under Section 504 may not be classified but do have a permanent or temporary impairment in a major life function (for example: performing manual tasks, walking, seeing, hearing, speaking, etc.).

Advanced planning is integral to implementing accommodations/modifications effectively and ensuring that the security of test materials is maintained. If a student requires an accommodation or modification that is not listed below, contact the Office of Evaluation and Assessment, GEPA Coordinator.

Accommodations must be recorded on the student's answer folder by the codes (A, B, C, or D) listed in this appendix. If using pre-ID labels, verify coding is correct.

ACCEPTABLE ACCOMMODATIONS OR MODIFICATIONS

Code

A. Setting Accommodations

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

4. Providing special lighting
5. Providing special furniture (e.g., desks, trays, carrels)

B. Scheduling Accommodations

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

C. Test Materials Modifications

1. Administering the large-print version of test materials
2. Administering the Braille version of test materials
3. Allowing separate additional continuation pages for writing tasks. These pages **MUST** be properly marked to link them to the correct student for credit.

D. Test Procedures Modifications

1. Administration modifications
 - a. reading directions aloud
 - b. reading test items aloud (**YOU MAY NOT READ ALOUD OR SIGN THE READING PASSAGES IN LANGUAGE ARTS LITERACY—YOU MAY READ ONLY THE READING ITEMS ASSOCIATED WITH THE PASSAGE**); **ONLY the teacher who must read test items aloud is permitted to have a test book assigned to them for this task.**
 - c. providing and ensuring that amplification (hearing aid and/or FM system) is in working order
 - d. using a sign language or cued speech interpreter for administration of directions or items **but not reading passages**
 - e. masking a portion of the test booklet and/or answer folder to eliminate visual distractors or providing reading windows
 - f. repeating, clarifying, or rewording directions
 - g. providing written directions on a separate sheet or transparency
 - h. using an examiner who is familiar with the student
 - i. using an examiner who can communicate fluently in sign language (American Sign Language or a form of Manually Coded English)
 - j. providing manipulatives for math items
 - k. using graph paper for math section
 - l. using a Braille ruler and talking calculator
 - m. using tactile or visual cues for deaf or hard of hearing students to indicate time to begin, time remaining, and time to end a particular part of the test

2. Response modifications
 - a. having an examiner record the student's identifying information on the answer folder, or grid corrections to the pre-ID label
 - b. dictating oral responses to a scribe (person who writes from dictation) – student must indicate all punctuation and must spell all key words (see FAQ – *Test Manual*, Appendix J)
 - c. using a Braille writer to record responses
 - d. signing responses to a sign language interpreter (student must indicate all punctuation and must spell all key words)
 - e. recording responses on a word processor
 - f. using large-face calculators
 - g. using talking calculators
 - h. providing an Augmentative Communication device
 - i. using a larger diameter or modified special grip #2 pencil
 - j. masking portions of the answer folder to eliminate visual distractors
 - k. marking answers in the test booklet (an examiner would transfer the answers to an answer folder)

OTHER CONSIDERATIONS

Ensure that:

- a. any medication has been appropriately adjusted so it will not interfere with the student's functioning.
- b. eyeglasses are used, if needed.
- c. hearing aids, FM systems, Augmentative Communication devices, word processors, or other equipment are functioning properly.
- d. source and strength of light are appropriate.
- e. all students can clearly see and hear the examiner.
- f. all deaf or hard of hearing students who communicate aurally/orally are watching the examiner when instructions are given.
- g. responses to open-ended items and writing tasks which are written or typed on separate sheets of paper by students eligible for this accommodation are labeled with student data paper-clipped to the front of the answer folder, and placed in the fluorescent orange envelope provided. Follow packaging instructions in this manual or the student's responses cannot be linked to their responses on the other sections of the test and they will receive incomplete scores. Copies of these pages should be made and retained on file by the school district until scores are received.

- h. students using the large-print test booklets
 1. mark their answers in the large-print answer folder.
 2. may be instructed to skip items identified in the LP instructions. The spaces for these items must be left blank on the student's answer folder (included in the large-print kit).
 3. who dictate responses on open-ended items and writing tasks indicate all punctuation and spell all key words.
- i. students using the Braille test booklets
 1. are instructed to bring a Braille ruler and a talking calculator to the test session.
 2. are instructed to skip dropped items identified in the Braille instructions. The spaces for these items must be left blank on the student transcription answer folder (included in the Braille kit).
 3. have answer folders transcribed from the Braille version by the Examiner.
 4. dictate their answers to the examiner or use a device that produces Braille. For dictations and responses recorded in Braille:
 - Students must indicate all punctuation and must spell all key words.
 - Examiners must transcribe the Braille responses into the regular answer folder included in the Braille kit.
- j. students who communicate in sign language
 1. have an interpreter to translate oral directions and test items (but not the Reading passages in the Language Arts Literacy section of the test). The interpreter should be able to communicate in the mode used by the student, American Sign Language or a form of Manually Coded English. The interpreter should be instructed to interpret so as not to give the answer to the student through the use of a particular sign or finger spelling.
 2. using American Sign Language for open-ended and writing task responses will sign the responses to the interpreter who will interpret them into spoken English and a scribe will record the responses in the answer folder.
 3. using Signed English or cued speech will sign/cue to the interpreter who will transliterate (word for word) into spoken English and a scribe will record the responses.

For any unresolved questions, contact the Office of Special Education Programs at (609) 292-2912.

Appendix D

Raw to Scale Scores Conversions

2006 GEPA LAL Raw Score to Scale Score

RS	Ability	SS
0.0	-5.6713	103
0.5	-5.0668	104
1.0	-4.4623	107
1.5	-4.1141	108
2.0	-3.7659	111
2.5	-3.5629	113
3.0	-3.3599	115
3.5	-3.2162	117
4.0	-3.0725	119
4.5	-2.9609	120
5.0	-2.8493	122
5.5	-2.7574	124
6.0	-2.6654	126
6.5	-2.5864	127
7.0	-2.5074	129
7.5	-2.4370	131
8.0	-2.3667	133
8.5	-2.3021	134
9.0	-2.2375	136
9.5	-2.1767	138
10.0	-2.1160	139
10.5	-2.0575	141
11.0	-1.9990	143
11.5	-1.9416	144
12.0	-1.8842	146
12.5	-1.8271	147
13.0	-1.7701	149
13.5	-1.7128	150
14.0	-1.6554	151
14.5	-1.5975	153
15.0	-1.5396	154
15.5	-1.4809	156
16.0	-1.4222	157
16.5	-1.3628	159
17.0	-1.3034	160
17.5	-1.2434	162
18.0	-1.1833	163
18.5	-1.1228	164
19.0	-1.0622	166
19.5	-1.0013	167
20.0	-0.9403	169
20.5	-0.8789	170
21.0	-0.8176	172

RS	Ability	SS
21.5	-0.7558	173
22.0	-0.6941	175
22.5	-0.6318	177
23.0	-0.5695	178
23.5	-0.5066	180
24.0	-0.4437	181
24.5	-0.3799	183
25.0	-0.3161	185
25.5	-0.2512	186
26.0	-0.1864	188
26.5	-0.1201	190
27.0	-0.0538	192
27.5	0.0143	193
28.0	0.0823	195
28.5	0.1524	197
29.0	0.2225	200*
29.5	0.2952	201
30.0	0.3678	203
30.5	0.4434	204
31.0	0.5190	206
31.5	0.5982	208
32.0	0.6773	210
32.5	0.7605	212
33.0	0.8437	214
33.5	0.9318	216
34.0	1.0198	218
34.5	1.1136	221
35.0	1.2074	223
35.5	1.3079	225
36.0	1.4085	227
36.5	1.5172	229
37.0	1.6260	231
37.5	1.7443	233
38.0	1.8627	236
38.5	1.9923	238
39.0	2.1218	240
39.5	2.2638	242
40.0	2.4057	245
40.5	2.5602	247
41.0	2.7147	250*
41.5	2.8805	252
42.0	3.0463	255
42.5	3.2209	257

RS	Ability	SS
43.0	3.3956	260
43.5	3.5767	263
44.0	3.7578	266
44.5	3.9449	268
45.0	4.1320	271
45.5	4.3269	274
46.0	4.5218	277
46.5	4.7270	280
47.0	4.9322	283
47.5	5.1484	285
48.0	5.3647	288
48.5	5.5909	290
49.0	5.8170	292
49.5	6.0543	294
50.0	6.2917	296
50.5	6.5492	298
51.0	6.8067	299
51.5	7.1108	300
52.0	7.4149	300
52.5	7.8514	300
53.0	8.2880	300
53.5	8.9470	300
54.0	9.6060	300

* Rounding was applied for this cut score.

2006 GEPA Mathematics Raw Score to Scale Score

RS	Ability	SS
0.0	-5.5502	137
0.5	-4.9306	137
1.0	-4.3109	138
1.5	-3.9401	140
2.0	-3.5694	141
2.5	-3.3434	142
3.0	-3.1174	143
3.5	-2.9512	144
4.0	-2.7849	145
4.5	-2.6515	146
5.0	-2.5181	148
5.5	-2.4057	149
6.0	-2.2932	150
6.5	-2.1952	151
7.0	-2.0973	152
7.5	-2.0100	153
8.0	-1.9226	155
8.5	-1.8434	156
9.0	-1.7642	157
9.5	-1.6915	158
10.0	-1.6187	159
10.5	-1.5510	161
11.0	-1.4834	162
11.5	-1.4201	163
12.0	-1.3567	164
12.5	-1.2968	166
13.0	-1.2370	167
13.5	-1.1801	168
14.0	-1.1232	169
14.5	-1.0688	171
15.0	-1.0145	172
15.5	-0.9622	173
16.0	-0.9100	174

RS	Ability	SS
16.5	-0.8595	176
17.0	-0.8091	177
17.5	-0.7602	178
18.0	-0.7113	180
18.5	-0.6637	181
19.0	-0.6162	182
19.5	-0.5697	184
20.0	-0.5232	185
20.5	-0.4776	187
21.0	-0.4320	188
21.5	-0.3871	189
22.0	-0.3423	191
22.5	-0.2980	192
23.0	-0.2537	194
23.5	-0.2098	195
24.0	-0.1659	197
24.5	-0.1222	198
25.0	-0.0786	200
25.5	-0.0350	201
26.0	0.0086	203
26.5	0.0522	204
27.0	0.0959	206
27.5	0.1398	208
28.0	0.1837	209
28.5	0.2281	211
29.0	0.2725	213
29.5	0.3176	214
30.0	0.3626	216
30.5	0.4086	218
31.0	0.4546	220
31.5	0.5017	221
32.0	0.5489	223
32.5	0.5975	225

RS	Ability	SS
33.0	0.6462	227
33.5	0.6967	229
34.0	0.7473	231
34.5	0.8001	233
35.0	0.8529	235
35.5	0.9086	237
36.0	0.9643	238
36.5	1.0234	240
37.0	1.0826	242
37.5	1.1460	245
38.0	1.2094	247
38.5	1.2781	250*
39.0	1.3469	251
39.5	1.4222	253
40.0	1.4976	255
40.5	1.5814	257
41.0	1.6653	259
41.5	1.7603	261
42.0	1.8552	263
42.5	1.9653	265
43.0	2.0753	267
43.5	2.2071	269
44.0	2.3388	271
44.5	2.5043	272
45.0	2.6698	274
45.5	2.8961	276
46.0	3.1224	278
46.5	3.4947	280
47.0	3.8670	281
47.5	4.4886	283
48.0	5.1102	286

* Rounding was applied for this cut score.

2006 GEPA Science Raw Score to Scale Score

RS	Ability	SS
0.0	- 5.302	126
0.5	- 4.687	126
1.0	- 4.071	128
1.5	- 3.705	130
2.0	- 3.339	132
2.5	- 3.117	133
3.0	- 2.895	135
3.5	- 2.732	137
4.0	- 2.569	138
4.5	- 2.439	140
5.0	- 2.308	142
5.5	- 2.198	143
6.0	- 2.088	145
6.5	- 1.992	147
7.0	- 1.895	148
7.5	- 1.810	150
8.0	- 1.724	152
8.5	- 1.646	154
9.0	- 1.568	155
9.5	- 1.497	157
10.0	- 1.425	159
10.5	- 1.359	160
11.0	- 1.292	163
11.5	- 1.229	164
12.0	- 1.167	166
12.5	- 1.108	167
13.0	- 1.049	169
13.5	- 0.993	170
14.0	- 0.936	172
14.5	- 0.883	174
15.0	- 0.829	176
15.5	- 0.777	178
16.0	- 0.726	179
16.5	- 0.676	181
17.0	- 0.627	183
17.5	- 0.578	184

RS	Ability	SS
18.0	- 0.530	186
18.5	- 0.483	188
19.0	- 0.437	189
19.5	- 0.391	191
20.0	- 0.345	192
20.5	- 0.301	194
21.0	- 0.256	196
21.5	- 0.212	197
22.0	- 0.169	200*
22.5	- 0.125	201
23.0	- 0.082	202
23.5	- 0.040	204
24.0	0.003	205
24.5	0.045	207
25.0	0.087	208
25.5	0.129	210
26.0	0.171	212
26.5	0.213	213
27.0	0.255	215
27.5	0.297	216
28.0	0.339	218
28.5	0.381	219
29.0	0.423	221
29.5	0.465	222
30.0	0.507	224
30.5	0.550	225
31.0	0.593	227
31.5	0.636	228
32.0	0.679	230
32.5	0.723	231
33.0	0.767	233
33.5	0.812	234
34.0	0.856	236
34.5	0.902	238
35.0	0.948	239
35.5	0.995	240

RS	Ability	SS
36.0	1.041	242
36.5	1.090	243
37.0	1.138	244
37.5	1.188	246
38.0	1.237	247
38.5	1.289	248
39.0	1.340	250
39.5	1.394	252
40.0	1.447	253
40.5	1.503	254
41.0	1.559	256
41.5	1.618	257
42.0	1.677	258
42.5	1.739	259
43.0	1.801	261
43.5	1.867	263
44.0	1.932	264
44.5	2.003	265
45.0	2.074	266
45.5	2.150	267
46.0	2.227	269
46.5	2.311	270
47.0	2.395	271
47.5	2.490	272
48.0	2.584	273
48.5	2.692	275
49.0	2.800	276
49.5	2.928	277
50.0	3.056	278
50.5	3.216	280
51.0	3.376	281
51.5	3.594	282
52.0	3.812	283
52.5	4.174	284
53.0	4.536	286
53.5	5.148	291
54.0	5.761	296

* Rounding was applied for this cut score.

2006 GEPA LAL Braille Raw Score to Scale Score

RS	Ability	SS
0.0	-5.5714	103
0.5	-4.9567	104
1.0	-4.3419	107
1.5	-3.9809	109
2.0	-3.6200	112
2.5	-3.4075	114
3.0	-3.1951	117
3.5	-3.0454	119
4.0	-2.8957	121
4.5	-2.7805	123
5.0	-2.6653	126
5.5	-2.5708	128
6.0	-2.4764	130
6.5	-2.3949	132
7.0	-2.3134	134
7.5	-2.2399	136
8.0	-2.1665	138
8.5	-2.0978	140
9.0	-2.0292	142
9.5	-1.9631	143
10.0	-1.8971	145
10.5	-1.8321	147
11.0	-1.7672	149
11.5	-1.7026	150
12.0	-1.6379	152
12.5	-1.5730	154
13.0	-1.5081	155
13.5	-1.4428	157
14.0	-1.3775	158
14.5	-1.3119	160
15.0	-1.2463	161
15.5	-1.1806	163
16.0	-1.1148	165
16.5	-1.0490	166
17.0	-0.9832	168
17.5	-0.9173	169
18.0	-0.8514	171
18.5	-0.7854	173

RS	Ability	SS
19.0	-0.7194	174
19.5	-0.6531	176
20.0	-0.5868	178
20.5	-0.5200	179
21.0	-0.4532	181
21.5	-0.3857	183
22.0	-0.3181	185
22.5	-0.2493	186
23.0	-0.1806	188
23.5	-0.1104	190
24.0	-0.0401	192
24.5	0.0321	194
25.0	0.1044	196
25.5	0.1791	198
26.0	0.2538	200
26.5	0.3315	202
27.0	0.4092	204
27.5	0.4905	206
28.0	0.5719	208
28.5	0.6575	210
29.0	0.7432	212
29.5	0.8341	214
30.0	0.9249	216
30.5	1.0220	218
31.0	1.1190	221
31.5	1.2236	223
32.0	1.3281	225
32.5	1.4418	228
33.0	1.5554	230
33.5	1.6801	232
34.0	1.8048	234
34.5	1.9428	237
35.0	2.0807	239
35.5	2.2338	242
36.0	2.3868	244
36.5	2.5560	247
37.0	2.7252	250
37.5	2.9096	252

RS	Ability	SS
38.0	3.0939	255
38.5	3.2910	258
39.0	3.4881	261
39.5	3.6957	265
40.0	3.9034	268
40.5	4.1228	271
41.0	4.3422	275
41.5	4.5775	278
42.0	4.8128	281
42.5	5.0666	284
43.0	5.3204	287
43.5	5.5907	290
44.0	5.8611	293
44.5	6.1506	295
45.0	6.4401	297
45.5	6.7714	299
46.0	7.1027	300
46.5	7.5613	300
47.0	8.0199	300
47.5	8.6916	300
48.0	9.3634	300

2006 GEPA Mathematics Braille Raw Score to Scale Score

RS	Ability	SS
0.0	- 5.4241	137
0.5	- 4.8003	137
1.0	- 4.1765	139
1.5	- 3.8002	140
2.0	- 3.4239	142
2.5	- 3.1928	143
3.0	- 2.9618	144
3.5	- 2.7908	145
4.0	- 2.6197	147
4.5	- 2.4818	148
5.0	- 2.3439	149
5.5	- 2.2269	151
6.0	- 2.1100	152
6.5	- 2.0076	153
7.0	- 1.9052	155
7.5	- 1.8133	156
8.0	- 1.7215	158
8.5	- 1.6376	159
9.0	- 1.5538	161
9.5	- 1.4762	162
10.0	- 1.3985	163
10.5	- 1.3259	165
11.0	- 1.2533	166
11.5	- 1.1847	168
12.0	- 1.1161	169
12.5	- 1.0509	171
13.0	- 0.9856	173

RS	Ability	SS
13.5	- 0.9232	174
14.0	- 0.8607	176
14.5	- 0.8006	177
15.0	- 0.7405	179
15.5	- 0.6824	181
16.0	- 0.6243	182
16.5	- 0.5678	184
17.0	- 0.5114	185
17.5	- 0.4563	187
18.0	- 0.4013	189
18.5	- 0.3473	191
19.0	- 0.2933	192
19.5	- 0.2402	194
20.0	- 0.1871	196
20.5	- 0.1344	198
21.0	- 0.0818	200
21.5	- 0.0294	201
22.0	0.0231	203
22.5	0.0757	205
23.0	0.1283	207
23.5	0.1815	209
24.0	0.2347	211
24.5	0.2888	213
25.0	0.3430	215
25.5	0.3987	217
26.0	0.4543	220
26.5	0.5120	222

RS	Ability	SS
27.0	0.5697	224
27.5	0.6299	226
28.0	0.6902	228
28.5	0.7538	231
29.0	0.8174	233
29.5	0.8851	236
30.0	0.9529	238
30.5	1.0257	241
31.0	1.0986	243
31.5	1.1779	246
32.0	1.2572	250*
32.5	1.3448	251
33.0	1.4323	253
33.5	1.5306	256
34.0	1.6288	258
34.5	1.7417	260
35.0	1.8545	263
35.5	1.9883	265
36.0	2.1221	267
36.5	2.2889	270
37.0	2.4556	272
37.5	2.6820	274
38.0	2.9083	276
38.5	3.2793	279
39.0	3.6502	280
39.5	4.2698	282
40.0	4.8893	285

* Rounding was applied for this cut score.

2006 GEPA LAL Breach 1 Raw Score to Scale Score

RS	Ability	SS
0.0	-5.4089	103
0.5	-4.8224	105
1.0	-4.2359	108
1.5	-3.9076	110
2.0	-3.5793	112
2.5	-3.3888	114
3.0	-3.1984	117
3.5	-3.0608	119
4.0	-2.9232	121
4.5	-2.8123	123
5.0	-2.7014	125
5.5	-2.6058	127
6.0	-2.5103	129
6.5	-2.4242	131
7.0	-2.3381	133
7.5	-2.2582	136
8.0	-2.1782	138
8.5	-2.1024	140
9.0	-2.0266	142
9.5	-1.9536	144
10.0	-1.8806	146
10.5	-1.8097	148
11.0	-1.7387	149
11.5	-1.6693	151
12.0	-1.5999	153
12.5	-1.5316	155
13.0	-1.4633	156
13.5	-1.3960	158
14.0	-1.3286	159
14.5	-1.2621	161
15.0	-1.1955	163
15.5	-1.1296	164
16.0	-1.0637	166
16.5	-0.9983	167
17.0	-0.9329	169
17.5	-0.8678	171
18.0	-0.8028	172
18.5	-0.7380	174

RS	Ability	SS
19.0	-0.6733	176
19.5	-0.6086	177
20.0	-0.5439	179
20.5	-0.4790	180
21.0	-0.4142	182
21.5	-0.3491	184
22.0	-0.2841	186
22.5	-0.2186	187
23.0	-0.1531	189
23.5	-0.0870	191
24.0	-0.0210	192
24.5	0.0456	194
25.0	0.1123	196
25.5	0.1796	198
26.0	0.2469	200*
26.5	0.3151	201
27.0	0.3832	203
27.5	0.4522	205
28.0	0.5212	206
28.5	0.5913	208
29.0	0.6615	210
29.5	0.7330	212
30.0	0.8045	213
30.5	0.8778	215
31.0	0.9510	217
31.5	1.0266	219
32.0	1.1022	220
32.5	1.1808	222
33.0	1.2594	224
33.5	1.3417	225
34.0	1.4240	227
34.5	1.5110	229
35.0	1.5979	231
35.5	1.6902	232
36.0	1.7825	234
36.5	1.8809	236
37.0	1.9793	238
37.5	2.0838	239

RS	Ability	SS
38.0	2.1884	241
38.5	2.2989	243
39.0	2.4094	245
39.5	2.5250	247
40.0	2.6407	248
40.5	2.7610	250
41.0	2.8814	252
41.5	3.0073	254
42.0	3.1331	256
42.5	3.2664	258
43.0	3.3997	260
43.5	3.5430	262
44.0	3.6864	265
44.5	3.8414	267
45.0	3.9963	269
45.5	4.1615	272
46.0	4.3266	274
46.5	4.4991	277
47.0	4.6716	279
47.5	4.8520	282
48.0	5.0325	284
48.5	5.2279	286
49.0	5.4234	288
49.5	5.6465	291
50.0	5.8696	293
50.5	6.1356	295
51.0	6.4017	297
51.5	6.7328	299
52.0	7.0638	300
52.5	7.5354	300
53.0	8.0069	300
53.5	8.6904	300
54.0	9.3739	300

* Rounding was applied for this cut score.

Appendix E

Scale Scores with Frequencies

2006 GEPA Language Arts Literacy Scale Scores with Frequencies

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students	Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
108	1	0.0	1	0.0	169	626	0.6	9,113	8.6
111	6	0.0	7	0.0	170	632	0.6	9,745	9.2
115	20	0.0	27	0.0	171	6	0.0	9,751	9.2
117	4	0.0	31	0.0	172	759	0.7	10,510	9.9
119	35	0.0	66	0.1	173	765	0.7	11,275	10.6
120	7	0.0	73	0.1	174	1	0.0	11,276	10.6
122	53	0.0	126	0.1	175	836	0.8	12,112	11.4
124	26	0.0	152	0.1	176	8	0.0	12,120	11.4
126	80	0.1	232	0.2	177	837	0.8	12,957	12.2
127	32	0.0	264	0.2	178	797	0.7	13,754	12.9
128	1	0.0	265	0.2	179	7	0.0	13,761	12.9
129	99	0.1	364	0.3	180	902	0.8	14,663	13.8
131	47	0.0	411	0.4	181	970	0.9	15,633	14.7
133	132	0.1	543	0.5	182	6	0.0	15,639	14.7
134	88	0.1	631	0.6	183	981	0.9	16,620	15.6
136	157	0.1	788	0.7	184	6	0.0	16,626	15.6
138	126	0.1	914	0.9	185	1,094	1.0	17,720	16.6
139	204	0.2	1,118	1.1	186	1,165	1.1	18,885	17.7
141	141	0.1	1,259	1.2	187	9	0.0	18,894	17.7
143	235	0.2	1,494	1.4	188	1,184	1.1	20,078	18.9
144	206	0.2	1,700	1.6	189	6	0.0	20,084	18.9
146	309	0.3	2,009	1.9	190	1,291	1.2	21,375	20.1
147	227	0.2	2,236	2.1	191	12	0.0	21,387	20.1
148	1	0.0	2,237	2.1	192	1,400	1.3	22,787	21.4
149	297	0.3	2,534	2.4	193	1,397	1.3	24,184	22.7
150	265	0.2	2,799	2.6	194	15	0.0	24,199	22.7
151	360	0.3	3,159	3.0	195	1,468	1.4	25,667	24.1
153	317	0.3	3,476	3.3	196	16	0.0	25,683	24.1
154	423	0.4	3,899	3.7	197	1,664	1.6	27,347	25.7
155	4	0.0	3,903	3.7	198	26	0.0	27,373	25.7
156	390	0.4	4,293	4.0	199	2	0.0	27,375	25.7
157	439	0.4	4,732	4.4	200	1,716	1.6	29,091	27.3
158	1	0.0	4,733	4.4	201	1,760	1.7	30,851	29.0
159	431	0.4	5,164	4.9	202	8	0.0	30,859	29.0
160	521	0.5	5,685	5.3	203	1,945	1.8	32,804	30.8
161	2	0.0	5,687	5.3	204	2,012	1.9	34,816	32.7
162	481	0.5	6,168	5.8	205	28	0.0	34,844	32.7
163	569	0.5	6,737	6.3	206	2,179	2.0	37,023	34.8
164	543	0.5	7,280	6.8	207	3	0.0	37,026	34.8
166	609	0.6	7,889	7.4	208	2,355	2.2	39,381	37.0
167	597	0.6	8,486	8.0	210	2,571	2.4	41,952	39.4
168	1	0.0	8,487	8.0	212	2,687	2.5	44,639	41.9

2006 GEPA Language Arts Literacy Scale Scores with Frequencies (continued)

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students	Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
213	24	0.0	44,663	42.0	250	1,946	1.8	99,304	93.3
214	2,868	2.7	47,531	44.7	251	1	0.0	99,305	93.3
215	22	0.0	47,553	44.7	252	1,656	1.6	100,961	94.8
216	3,132	2.9	50,685	47.6	253	2	0.0	100,963	94.8
217	18	0.0	50,703	47.6	254	10	0.0	100,973	94.9
218	3,234	3.0	53,937	50.7	255	1,305	1.2	102,278	96.1
219	32	0.0	53,969	50.7	256	9	0.0	102,287	96.1
220	33	0.0	54,002	50.7	257	1,040	1.0	103,327	97.1
221	3,349	3.1	57,351	53.9	258	3	0.0	103,330	97.1
222	24	0.0	57,375	53.9	260	835	0.8	104,165	97.9
223	3,446	3.2	60,821	57.1	262	6	0.0	104,171	97.9
224	28	0.0	60,849	57.2	263	634	0.6	104,805	98.5
225	3,635	3.4	64,484	60.6	265	1	0.0	104,806	98.5
227	3,839	3.6	68,323	64.2	266	489	0.5	105,295	98.9
229	3,816	3.6	72,139	67.8	267	1	0.0	105,296	98.9
231	3,787	3.6	75,926	71.3	268	341	0.3	105,637	99.2
232	23	0.0	75,949	71.3	269	4	0.0	105,641	99.2
233	3,480	3.3	79,429	74.6	271	252	0.2	105,893	99.5
234	25	0.0	79,454	74.6	274	177	0.2	106,070	99.6
235	1	0.0	79,455	74.6	277	149	0.1	106,219	99.8
236	3,643	3.4	83,098	78.1	280	80	0.1	106,299	99.9
237	4	0.0	83,102	78.1	282	1	0.0	106,300	99.9
238	3,336	3.1	86,438	81.2	283	56	0.1	106,356	99.9
239	15	0.0	86,453	81.2	285	30	0.0	106,386	99.9
240	3,132	2.9	89,585	84.2	286	1	0.0	106,387	99.9
241	25	0.0	89,610	84.2	288	24	0.0	106,411	100.0
242	2,893	2.7	92,503	86.9	290	13	0.0	106,424	100.0
243	19	0.0	92,522	86.9	292	12	0.0	106,436	100.0
244	1	0.0	92,523	86.9	294	3	0.0	106,439	100.0
245	2,620	2.5	95,143	89.4	296	6	0.0	106,445	100.0
247	2,205	2.1	97,348	91.5	298	1	0.0	106,446	100.0
248	10	0.0	97,358	91.5	300	1	0.0	106,447	100.0

N-COUNT = 106,447 MEAN = 214.2940 STANDARD DEVIATION = 28.4738 SEM = 11.864

2006 GEPA Mathematics Scale Scores with Frequencies

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students	Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
138	13	0.0	13	0.0	186	2	0.0	26,437	24.6
141	31	0.0	44	0.0	187	1,109	1.0	27,546	25.6
143	71	0.1	115	0.1	188	1,452	1.4	28,998	27.0
144	12	0.0	127	0.1	189	1,094	1.0	30,092	28.0
145	187	0.2	314	0.3	190	1	0.0	30,093	28.0
146	42	0.0	356	0.3	191	1,446	1.3	31,539	29.3
148	376	0.3	732	0.7	192	1,171	1.1	32,710	30.4
149	83	0.1	815	0.8	194	1,496	1.4	34,206	31.8
150	585	0.5	1,400	1.3	195	1,235	1.1	35,441	33.0
151	165	0.2	1,565	1.5	196	4	0.0	35,445	33.0
152	811	0.8	2,376	2.2	197	1,501	1.4	36,946	34.4
153	233	0.2	2,609	2.4	198	1,273	1.2	38,219	35.5
154	1	0.0	2,610	2.4	199	1	0.0	38,220	35.5
155	996	0.9	3,606	3.4	200	1,801	1.7	40,021	37.2
156	336	0.3	3,942	3.7	201	1,397	1.3	41,418	38.5
157	1,099	1.0	5,041	4.7	203	1,724	1.6	43,142	40.1
158	408	0.4	5,449	5.1	204	1,354	1.3	44,496	41.4
159	1,077	1.0	6,526	6.1	206	1,836	1.7	46,332	43.1
160	4	0.0	6,530	6.1	208	1,438	1.3	47,770	44.4
161	506	0.5	7,036	6.5	209	1,820	1.7	49,590	46.1
162	1,153	1.1	8,189	7.6	210	2	0.0	49,592	46.1
163	572	0.5	8,761	8.1	211	1,498	1.4	51,090	47.5
164	1,157	1.1	9,918	9.2	213	1,853	1.7	52,943	49.2
165	1	0.0	9,919	9.2	214	1,524	1.4	54,467	50.7
166	646	0.6	10,565	9.8	215	1	0.0	54,468	50.7
167	1,168	1.1	11,733	10.9	216	1,897	1.8	56,365	52.4
168	659	0.6	12,392	11.5	218	1,548	1.4	57,913	53.9
169	1,147	1.1	13,539	12.6	220	1,993	1.9	59,906	55.7
170	1	0.0	13,540	12.6	221	1,569	1.5	61,475	57.2
171	729	0.7	14,269	13.3	223	1,964	1.8	63,439	59.0
172	1,200	1.1	15,469	14.4	225	1,574	1.5	65,013	60.5
173	777	0.7	16,246	15.1	227	1,950	1.8	66,963	62.3
174	1,205	1.1	17,451	16.2	229	1,597	1.5	68,560	63.8
176	817	0.8	18,268	17.0	231	2,041	1.9	70,601	65.7
177	1,223	1.1	19,491	18.1	232	2	0.0	70,603	65.7
178	943	0.9	20,434	19.0	233	1,566	1.5	72,169	67.1
180	1,299	1.2	21,733	20.2	234	1	0.0	72,170	67.1
181	984	0.9	22,717	21.1	235	2,083	1.9	74,253	69.1
182	1,346	1.3	24,063	22.4	236	1	0.0	74,254	69.1
183	2	0.0	24,065	22.4	237	1,563	1.5	75,817	70.5
184	978	0.9	25,043	23.3	238	2,088	1.9	77,905	72.4
185	1,392	1.3	26,435	24.6	240	1,515	1.4	79,420	73.9

2006 GEPA Mathematics Scale Scores with Frequencies (continued)

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
242	2,123	2.0	81,543	75.8
245	1,456	1.4	82,999	77.2
246	1	0.0	83,000	77.2
247	2,065	1.9	85,065	79.1
250	1,430	1.3	86,495	80.4
251	1,990	1.9	88,485	82.3
253	1,257	1.2	89,742	83.5
255	1,930	1.8	91,672	85.3
257	1,243	1.2	92,915	86.4
258	1	0.0	92,916	86.4
259	1,879	1.7	94,795	88.2
260	1	0.0	94,796	88.2
261	1,222	1.1	96,018	89.3
262	1	0.0	96,019	89.3
263	1,729	1.6	97,748	90.9
265	1,021	0.9	98,769	91.9
266	2	0.0	98,771	91.9
267	1,662	1.5	100,433	93.4
269	840	0.8	101,273	94.2
271	1,509	1.4	102,782	95.6
272	736	0.7	103,518	96.3
274	1,271	1.2	104,789	97.5
276	511	0.5	105,300	97.9
278	1,006	0.9	106,306	98.9
280	279	0.3	106,585	99.1
281	594	0.6	107,179	99.7
283	105	0.1	107,284	99.8
286	246	0.2	107,530	100.0

N-COUNT = 107,530 MEAN = 214.6122 STANDARD DEVIATION = 34.9881 SEM = 12.860

2006 GEPA Science Scale Scores with Frequencies

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students	Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
128	2	0.0	2	0.0	192	2,037	1.9	18,888	17.6
132	1	0.0	3	0.0	194	846	0.8	19,734	18.3
135	5	0.0	8	0.0	195	1	0.0	19,735	18.3
138	5	0.0	13	0.0	196	1,983	1.8	21,718	20.2
142	22	0.0	35	0.0	197	920	0.9	22,638	21.0
143	3	0.0	38	0.0	200	2,195	2.0	24,833	23.1
145	47	0.0	85	0.1	201	1,051	1.0	25,884	24.1
147	2	0.0	87	0.1	202	2,136	2.0	28,020	26.0
148	91	0.1	178	0.2	204	1,148	1.1	29,168	27.1
150	10	0.0	188	0.2	205	2,228	2.1	31,396	29.2
151	2	0.0	190	0.2	207	1,129	1.0	32,525	30.2
152	152	0.1	342	0.3	208	2,249	2.1	34,774	32.3
154	16	0.0	358	0.3	210	1,268	1.2	36,042	33.5
155	266	0.2	624	0.6	212	2,309	2.1	38,351	35.6
156	1	0.0	625	0.6	213	1,248	1.2	39,599	36.8
157	38	0.0	663	0.6	214	4	0.0	39,603	36.8
158	1	0.0	664	0.6	215	2,332	2.2	41,935	39.0
159	443	0.4	1,107	1.0	216	1,371	1.3	43,306	40.2
160	71	0.1	1,178	1.1	218	2,453	2.3	45,759	42.5
162	2	0.0	1,180	1.1	219	1,438	1.3	47,197	43.9
163	648	0.6	1,828	1.7	220	3	0.0	47,200	43.9
164	103	0.1	1,931	1.8	221	2,320	2.2	49,520	46.0
166	789	0.7	2,720	2.5	222	1,506	1.4	51,026	47.4
167	151	0.1	2,871	2.7	223	1	0.0	51,027	47.4
169	1,020	0.9	3,891	3.6	224	2,423	2.3	53,450	49.7
170	247	0.2	4,138	3.8	225	1,574	1.5	55,024	51.1
171	1	0.0	4,139	3.8	226	1	0.0	55,025	51.1
172	1,172	1.1	5,311	4.9	227	2,473	2.3	57,498	53.4
174	333	0.3	5,644	5.2	228	1,567	1.5	59,065	54.9
176	1,427	1.3	7,071	6.6	229	1	0.0	59,066	54.9
178	363	0.3	7,434	6.9	230	2,308	2.1	61,374	57.0
179	1,588	1.5	9,022	8.4	231	1,680	1.6	63,054	58.6
180	5	0.0	9,027	8.4	233	2,392	2.2	65,446	60.8
181	518	0.5	9,545	8.9	234	1,585	1.5	67,031	62.3
183	1,755	1.6	11,300	10.5	235	3	0.0	67,034	62.3
184	573	0.5	11,873	11.0	236	2,278	2.1	69,312	64.4
186	1,761	1.6	13,634	12.7	238	1,595	1.5	70,907	65.9
187	4	0.0	13,638	12.7	239	2,291	2.1	73,198	68.0
188	661	0.6	14,299	13.3	240	1,593	1.5	74,791	69.5
189	1,816	1.7	16,115	15.0	241	1	0.0	74,792	69.5
190	1	0.0	16,116	15.0	242	2,182	2.0	76,974	71.5
191	735	0.7	16,851	15.7	243	1,559	1.4	78,533	73.0

2006 GEPA Science Scale Scores with Frequencies (continued)

Scale Score	Number of Students	Percent of Students	Cumulative Number of Students	Cumulative Percent of Students
244	2,058	1.9	80,591	74.9
245	3	0.0	80,594	74.9
246	1,551	1.4	82,145	76.3
247	1,948	1.8	84,093	78.2
248	1,484	1.4	85,577	79.5
250	1,766	1.6	87,343	81.2
252	1,370	1.3	88,713	82.4
253	1,751	1.6	90,464	84.1
254	1,268	1.2	91,732	85.3
255	3	0.0	91,735	85.3
256	1,609	1.5	93,344	86.8
257	1,176	1.1	94,520	87.8
258	1,400	1.3	95,920	89.1
259	1,118	1.0	97,038	90.2
261	1,229	1.1	98,267	91.3
263	915	0.9	99,182	92.2
264	1,091	1.0	100,273	93.2
265	835	0.8	101,108	94.0
266	923	0.9	102,031	94.8
267	679	0.6	102,710	95.5
269	761	0.7	103,471	96.2
270	591	0.5	104,062	96.7
271	676	0.6	104,738	97.3
272	451	0.4	105,189	97.8
273	491	0.5	105,680	98.2
275	368	0.3	106,048	98.6
276	393	0.4	106,441	98.9
277	275	0.3	106,716	99.2
278	296	0.3	107,012	99.5
280	156	0.1	107,168	99.6
281	174	0.2	107,342	99.8
282	83	0.1	107,425	99.8
283	89	0.1	107,514	99.9
284	36	0.0	107,550	100.0
286	38	0.0	107,588	100.0
291	7	0.0	107,595	100.0
296	5	0.0	107,600	100.0

N-COUNT = 107,600 MEAN = 223.1891 STANDARD DEVIATION = 28.5999 SEM = 10.891

Appendix F

Reporting Limited English Proficiency (LEP) and Special Education (SE)

LIMITED ENGLISH PROFICIENT (LEP)

A limited English proficient student is a student whose native language is one other than English. This student has sufficient difficulty speaking, reading, writing, or understanding the English language, as measured by an English language proficiency test, so as to be denied the opportunity to learn successfully in the classroom where the language of instruction is English.

School staff were instructed to mark a circle to designate the number of academic years each limited English proficient student participated in a language assistance program (Bilingual, English as a Second Language, or English Language Services) in **ANY** school in their **DISTRICT**. The codes for LEP are:

- < = LEP student **entered** a language assistance program **AFTER July 1, 2005**, and is **currently enrolled in the program**. *These students do not have to take the LAL portion of the test but MUST take Math and Science.*
- 1 = LEP student **entered** a language assistance program **BETWEEN July 1, 2004, and June 30, 2005**, and is **currently enrolled in the program**.
- 2 = LEP student **entered** a language assistance program **BETWEEN July 1, 2003, and June 30, 2004**, and is **currently enrolled in the program**.
- 3 = LEP student **entered** a language assistance program **BEFORE July 1, 2003**, and is **currently enrolled in the program**.
- F = **Former** LEP student **exited** a language assistance program **AFTER July 1, 2003, and is NO longer enrolled in the program**.

LIMITED-ENGLISH PROFICIENT (LEP) EXEMPT – LAL

- E = LEP student **entered the United States** as well as a language assistance program **AFTER July 1, 2005 [currently enrolled in the program]**. *These students do not have to take the LAL portion of the test but MUST take Math and Science.*

SPECIAL EDUCATION (SE)

There are 13 codes for Special Education categories. The categories are:

- A. Auditorily Impaired
- B. Other Health Impaired
- C. Communication Impaired
- D. Emotionally Disturbed
- E. Cognitively Impaired
- F. Multiply Disabled
- G. Traumatic Brain Injury
- H. Orthopedically Impaired
- I. Specific Learning Disability
- J. Social Maladjustment
- K. Visually Impaired
- L. Speech-Language Services Only
- M. Autistic

For reporting, category N is used to indicate multiple grids. This is also a default code when a school failed to provide the specific information listed above for an APA student.

References

REFERENCES

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1974). *Standards for Educational and Psychological Tests*. Washington, DC: American Psychological Association.

American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

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

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.

Crocker, L. & Algina, J. (1986). *Introduction to Classical & Modern Test Theory*. Belmont, CA: Wadsworth Group.

Cycle II Criterion-Based Holistic Scoring: A Mathematics and Science Handbook. Developed by the New Jersey Department of Education, January 2006.

Cycle II Criterion-Based Holistic Scoring: A Reading Handbook. Developed by the New Jersey Department of Education, January 2004.

Cycle II Criterion-Based Holistic Scoring: A Writing Handbook. Developed by the New Jersey Department of Education, September 2006.

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), 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 by the New Jersey Department of Education, February 1998.

Directory of Test Specifications and Sample Items for the Grade Eight Proficiency Assessment (GEPA) and High School Proficiency Assessment (HSPA) in Science. Developed by the New Jersey Department of Education, February 1998.

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.

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 that annual meeting of the American Educational Research Association, San Diego, CA.
- Kane, M. T. (2006). Validation. In R.L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 17-64). Westport, CT: American Council on Education and Praeger.
- Kolen, M.J. & Brennan, R.L. (2004). *Test Equating, Scaling and Linking*. New York: Springer Science and Business Media, Inc.
- Linacre, J. M. (2006) *A User's Guide to WINSTEPS MINISTEP Rasch-Model Computer Programs*. Chicago: Winsteps.com
- Linacre, J. M. (2006) *WINSTEPS Rasch measurement computer program*. Chicago: Winsteps.com
- Livingston, S.A. & Zieky, M.J. (1982) *Passing Scores*. Princeton, NJ: Educational Testing Service.
- Masters, G. N & Wright, B. (1997). The partial credit model. In W. J. van der Linden & R.K. Hambleton, (Eds.), *Handbook of Modern Item Response Theory*. New York. Springer-Verlag.
- Mehrens, W.A. & Lehmann, I.J. (1991) *Measurement and Evaluation in Education and Psychology* (4th ed.). New York: Holt, Rinehart and Winston.
- Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35, 1012-1027.
- Messick, S. (1995) Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice* 14(4), 5-8.
- Millman, J. & Greene, J. (1989) The specification and development of tests of achievement and ability. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 335-366) New York: American Council on Education.
- New Jersey Core Curriculum Content Standards*. Developed by the New Jersey Department of Education, October 2004. Retrieved January 4, 2007, from <http://www.state.nj.us/njded/cccs/cccs.pdf>
- Score Interpretation Manual*. Developed by the New Jersey Department of Education, June 2006.
- Van Der Linden, W.J. & Hambleton, R.K. (Eds.)(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.
- Zhao, H., Robinson, B., & Guo, F. (2006) *How well the New Jersey's grade eight proficiency assessment serves its designed purpose*. 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.



PTM#1506.20

Copyright © 2007 New Jersey

Department of Education

All rights reserved.