SOUTH CAROLINA
END-OF-COURSE EXAMINATION PROGRAM

2003–04 OPERATIONAL TEST AND IMPLEMENTATION TEST TECHNICAL REPORT

Prepared by

Bokhee Yoon, Kyunghee Suh, and Karen Thornton

American Institutes for Research

Edited and Issued by the

South Carolina Department of Education

Inez M. Tenenbaum
State Superintendent of Education

March 2006

CONTENTS

CHAPTER 1. HISTORY AND OVERVIEW 6
1.1 Planning Meeting 8
1.2 Passage Review Meeting 8
1.3 Development of Measurement Guidelines and Item Specifications 8
1.4 Development of Field-Test Item Pool 11
1.5 External Item Review 11
1.6 Development of Field-Test Forms 13

CHAPTER 2. STUDENT DEMOGRAPHICS 16
2.1 Student Participation 16
2.2 Accommodations 17
2.3 Test Administration Time 18
2.4 Student Questionnaire 20

CHAPTER 3. TEST ADMINISTRATION 21
3.1 Test Administration Window 21
3.2 Timing of the Test 21
3.3 Administration Manuals 21
3.4 Customized Materials 22
3.5 Pretest Workshops and Training 23
3.6 Materials Shipping and Return 24

CHAPTER 4. TECHNICAL CHARACTERISTICS OF ITEMS 25
4.1 Item Nonresponse Rates 25
4.2 Classical Item Statistics 25

CHAPTER 5. ITEM CALIBRATION AND SCALING 27
5.1 Methodology and Software 27
5.2 Pre-Equating 27
5.3 Item Calibration 27
5.4 Scaling 28
5.5 Definition of Scoreability 28
5.6 Reporting of Zero and Perfect Scores 28
5.7 Cut Scores for Passing Scores 29
5.8 Percentage of Students Scoring in Each Letter-Grade Equivalent 30

CHAPTER 6. DESCRIPTIVE STATISTICS 42

CHAPTER 7. RELIABILITY 48
7.1 Reliability of Raw Scores 48
7.2 Overall and Conditional SEM 48
7.3 Consistency of Passing Cut Scores 49

CHAPTER 8. VALIDITY 51
8.1 Item Distribution across Content Domains 51
8.2 Item Development 53
8.3 Differential Item Functioning 53
8.4 Correlations among Content Domains 56

WORKS CITED 58

LISTS OF TABLES AND FIGURES

TABLES

1.1 English 1 Test Blueprint 13

1.2 Biology 1/Applied Biology 2 Test Blueprint.....13

1.3 Physical Science Test Blueprint.....14

2.1 Summary of Student Demographics in the Sample 16

2.2 Accommodations Used in 2003–04 EOCEP Testing 18

2.3 Time Taken in 2003–04 EOCEP Testing with Regular Forms 19

2.4 Time Taken in 2003–04 EOCEP Testing with Customized Forms 20

4.1 Summary of Classical Item Statistics 25

5.1 Cut Scores on the Rasch Ability Scale and the Reporting Scale 29

5.2 Algebra 1/Math Tech 2 Operational Test, Regular Schools:
Percentages of Student Scores in Letter-Grade Equivalents,
Overall and by Demographics 30

5.3 Algebra 1/Math Tech 2 Operational Test, Adult Education Programs:
Percentages of Student Scores in Letter-Grade Equivalents,
Overall and by Demographics 32

5.4 English 1 Implementation Test, Regular Schools:
Percentages of Student Scores in Letter-Grade Equivalents,
Overall and by Demographics 33

5.5 English 1 Implementation Test, Adult Education Programs:
Percentages of Student Scores in Letter-Grade Equivalents,
Overall and by Demographics 34

5.6 Biology 1/Applied Biology 2 Implementation Test, Regular Schools:
Percentages of Student Scores in Letter-Grade Equivalents,
Overall and by Demographics 35

5.7 Biology 1/Applied Biology 2 Implementation Test, Adult Education Programs:
Percentages of Student Scores in Letter-Grade Equivalents,
Overall and by Demographics 37

5.8 Physical Science Implementation Test, Regular Schools:
Percentages of Student Scores in Letter-Grade Equivalents,
Overall and by Demographics 39

5.9 Physical Science Implementation Test, Adult Education Programs:
Percentages of Student Scores in Letter-Grade Equivalents,
Overall and by Demographics 40

6.1 2003–04 EOCEP Test Administration Summary Statistics:
Regular Schools and Adult Education Programs,
Overall and by Gender and Race 42

6.2 Summary Statistics: Algebra 1/Math Tech 2 Scrambled Forms 48
7.1 Reliability Coefficients of Raw Scores 48
7.2 Conditional Standard Errors of Measurement 49
7.3 Consistency Index for Passing Scores 50
8.1 Item Distribution by Content Domain for Algebra 1/Math Tech 2 51
8.2 Item Distribution by Content Domain for English 1 52
8.3 Item Distribution by Content Domain for Biology 1/Applied Biology 2 52
8.4 Item Distribution by Content Domain for Physical Science 53
8.5 Summary of Differential Item Functioning for Operational Items 55
8.6 Summary of Differential Item Functioning for Field-Test Items 56
8.7 Correlations among Domain Scores for Algebra 1/Math Tech 2 56
8.8 Correlations among Domain Scores for English 1 57
8.9 Correlations among Domain Scores for Biology 1/Applied Biology 2 and Physical Science 57

FIGURES

6.1 Algebra 1/Mathematics for the Technologies 2 (Form 30) 43
6.2 Algebra 1/Mathematics for the Technologies 2 (Form 40) 44
6.3 Algebra 1/Mathematics for the Technologies 2 (Form 10) 44
6.4 English 1(Form 10) 45
6.5 English 1(Form 20) 45
6.6 Biology 1/Applied Biology 2 (Form 10) 46
6.7 Biology 1/Applied Biology 2 (Form 40) 46
6.8 Physical Science (Form 10) 47
6.9 Physical Science (Form 20) 47

Chapter 1

HISTORY AND OVERVIEW

The South Carolina Education Accountability Act of 1998 requires the development of end-of-course examinations in gateway, or benchmark, courses for grades nine through twelve. When the program is fully implemented, all students enrolled in End-of-Course Examination Program (EOCEP) courses will take the tests for those courses: Algebra 1, Mathematics for the Technologies 2, English 1, Physical Science, Biology 1, Applied Biology 2, and U.S. History and Constitution.

As they are enunciated in State Board of Education Regulation 43-262.4, the purposes and uses of the EOCEP tests are as follows:

A. The tests shall promote instruction in the specific academic standards for the courses, encourage student achievement, and document the level of students' mastery of the curriculum standards.

B. The tests shall serve as indicators of program, school, and school district effectiveness in the manner prescribed by the Education Oversight Committee in accordance with the provisions of the Education Accountability Act of 1998 (EAA).

C. The tests shall be weighted 20 percent in the determination of students' final grades in the gateway courses.

EOCEP exams will be reported on the basis of the South Carolina uniform grading scale. The score reported is a scale score and not the percentage of correct answers.

The Algebra 1/Mathematics for the Technologies 2 end-of-course examination was implemented in the baseline year 2002–03 and was operational for the first time in 2003–04. The English 1, Physical Science, and Biology 1/Applied Biology 2 examinations that were field-tested in May 2003 were implemented for the baseline year in 2003–04. These subject-area EOCEP examinations will become operational in 2004–05. The U.S. History and Constitution examination is scheduled for field-testing in 2005–06, with baseline implementation in 2006–07 and operational administrations beginning in 2007–08.

The State Department of Education (SDE) awarded the contract for the development and scoring of the EOCEP tests in October 2001 to American Institutes for Research (AIR) and its partners Insite, Inc., and Pearson Educational Measurement (PEM). These contractors have undertaken a number of development, review, implementation, and data analysis activities.

The AIR conducted a field test in spring 2003 to create a precalibrated item pool with a sufficient number of items to construct the pre-equated operational-test forms for English 1, Biology 1/Applied Biology 2, and Physical Science tests. Writing original multiple-choice items rather than selecting and/or adapting items from an existing item pool is the approach that was considered most efficient and effective for achieving this goal. Item development for English 1, Biology 1/Applied Biology 2, and Physical Science was carried out during the 2002–03 year. The second half of item development for the Algebra 1/Mathematics for the Technologies 2

examination was completed during 2002–03 according to item and test specifications set in 2001–02.

The fact that the SDE involved South Carolina educators in making decisions about measuring the state's curriculum standards necessitated two types of committees regarding the field-test development process: a content review committee (CRC) in each subject area and a sensitivity review committee (SRC). The CRC in each subject area comprised expert representatives from three groups: SDE curriculum and assessment staff, district-level curriculum staff, and higher education faculty members. The SRC was formed by inviting individuals from organizations such as the Commission for Minority Affairs and the Association for Rural Education, as well as representatives from women's study programs, people with disabilities, and school counselors. Several of the committee members had served on previous SDE bias and sensitivity review committees for test development projects. These committees were integral to ensuring bias-free items and the alignment of test items with the standards in the test-development process. The roles of the committees in defining the standards, reviewing field-test passages and items, and developing the blueprints for the examinations are discussed later in this chapter.

The following steps were taken in developing the EOCEP tests for English 1, Biology 1/Applied Biology 2, and Physical Science:

1. An initial planning meeting for these three EOCEP tests was held in August 2002 with subject-area staff from the SDE's Office of Assessment and the Office of Curriculum and Standards and subject-area CRC members. At these meetings, draft measurement guidelines and taxonomies for each subject were reviewed, subject-specific measurement issues were identified, and preliminary item and passage writing specifications were set.

2. An SRC meeting was convened in September 2002 to review potential passages to be used on the English 1 examination. It was important that the proposed passages be reviewed and approved by the SRC before item writing could begin for English 1.

3. Preliminary measurement guidelines and item specifications were reviewed and revised by the partial CRCs at the August 2002 meeting. Review, revision, and approval of the measurement guidelines and item specifications—which were to be used as guides in the creation of the test items—occurred at the full CRC meetings held in October 2002. At the October meetings, the committees allocated by domain the item percentages comprising the examinations. This information formed the basis for the blueprint that was created for each end-of-course examination.

4. Development of the field-test item pools began immediately following the October 2002 CRC meetings. Qualified item writers were trained to write items specifically aligned with the South Carolina curriculum standards for each subject and to write in the proportions as identified by the test blueprints. Approximately 3,000 items were developed and processed through numerous AIR and Insite reviews for content validity, technical merit, and bias/sensitivity issues. After extensive internal review, the items were submitted to the SDE for review.

5. The CRCs reconvened in January 2003 to review the field-test items and to finalize the measurement guidelines and test blueprints. The SRC met immediately following the CRC meetings to review the items for possible bias or sensitivity issues. AIR and Insite content

specialists reviewed all comments, discussed possible item edits with SDE staff, and made suitable revisions.

6. On the basis of the approved test blueprints, twelve field-test forms containing 60 items each, including 15 anchor items, were constructed for the Biology 1/Applied Biology 2 and Physical Science examinations. Fourteen field-test forms of 60 items each, including 15 anchor items, were created for the English 1 examination. For the spring 2003 examination, 199 field-test items for Algebra 1/Mathematics for the Technologies 2 were incorporated into twenty forms.

1.1 PLANNING MEETING

The primary purpose of the first CRC meetings, held in August 2002, was to assist in the establishment of frameworks for the EOCEP examinations. To prepare for the upcoming meetings in October, AIR and Insite staff identified general and standard-specific measurement issues in each of the three content areas and presented them to the SDE for discussion. These same issues were then presented as topics for discussion at October 2002 meetings so that thorough CRC review and input could be achieved.

1.2 PASSAGE REVIEW MEETING

Before item writing for the English 1 examination was begun, the reading passages for which the items would be written had to be reviewed by the SRC to identify any passages that contained language or subject matter that could be potentially offensive to a particular subgroup of students and/or that could give a particular subgroup a performance advantage. SRC participants individually reviewed the passages, recorded notes on the passages, and flagged any problematic passages. The group then identified and discussed issues and reached agreement on appropriate revisions. SRC revisions and recommendations were documented so that passages could be either suitably edited or deleted.

1.3 DEVELOPMENT OF MEASUREMENT GUIDELINES AND ITEM SPECIFICATIONS

As part of the item and test development process, measurement guidelines and item specifications were drafted to ensure that the tests were aligned with the curriculum standards. Rather than using predetermined test specifications, CRC members met in the summer and fall of 2002 to generate measurement guidelines and item specification standards through a definition and clarification process. These guidelines and standards included documentation concerning decisions about general testing issues (e.g., use of calculators, ways to test “inquiry” in the sciences); assignment of preliminary item allocations; test blueprint and form construction plans; standard-specific issues (i.e., what could be covered on a multiple-choice test and what was more suitable for classroom assessment); and item format and content specifications.

Preliminary measurement guidelines and item specification standards were identified by a subset of CRC members and SDE staff during planning meetings held in August 2002. At the October 2002 meetings, the full CRCs identified the standards that would be tested by the EOCEP

examinations (some standards are not suitable for assessment with multiple-choice items), refined the measurement guidelines and the item specifications, and determined the preliminary allocation of items by domain and subdomain. Using a sample of test items specifically aligned with the standards to keep their minds focused on the task at hand, CRC members discussed general and item-specific issues. During their meetings in January 2003, after reviewing the proposed field-test items, the CRCs finalized the measurement guidelines and item specifications.

The test blueprints used in the construction of test forms were based on CRC judgments of the percentages of items that should be allocated to the various domains and subdomains on each EOCEP examination. The preliminary allocations were made by the committee members at their October 2002 meetings and were reviewed and finalized at their January 2003 meetings.

The following is a summary of specific issues the AIR and Insite staff addressed with the CRCs at the October meetings:

- Use of calculators for Physical Science

  In an August meeting between SDE staff and a subset of CRC members, it had been recommended that calculators be allowed during the EOCEP Physical Science test. At the full committee meeting in October, the CRC members decided against allowing the use of calculators and requested that any problems on the test be constructed so that students could solve them without the use of a calculator.

- Length of the test

  The CRCs were informed of the proposed test length of 60 items, with the understanding that this number was approximate and that as the items were developed and the tests were assembled, the test length might change. Nevertheless, the items for each EOCEP exam would cover the state standards and would meet the reliability and time requirements stipulated in the SDE's request for proposals (that is, the test length would allow 90 percent of the students to complete the test in 100 minutes.)

- Physical Science reference materials

  An equation reference sheet and the periodic table of elements were presented to the Physical Science CRC for review. Members were asked if they felt students should be given equation reference sheets and the periodic table of elements to use as reference materials during the examination, and if so, which equations should be presented and what level of detail the periodic table should cover. These materials were made available to students during the field test, and the full committee strongly urged that they be made available at the beginning of each Physical Science course so that students could become familiar with them prior to testing.

- Use of graphics in the Biology 1/Applied Biology 2 and Physical Science examinations

  The CRCs requested that a variety of graphics (tables, charts, graphs, experiment setups, etc.) be used on each of these EOCEP exams.

- Whether the Communications and Research standards in English 1 should be tested and, if so, how they should be tested

The October meetings were also devoted to clarifying measurement guidelines and further defining the curriculum standards to be tested so that CRC participants could provide guidance for subsequent item-development work. The meetings included in-depth discussion of the standards and related issues, review of standards with sample item types, and examination of item allocations by domain.

As a first activity in further clarifying the standards, CRC members discussed each standard in their particular subject area and considered its appropriateness for a multiple-choice end-of-course examination. At the end of this activity, each committee member made a preliminary judgment about how to allocate items to domains by first determining the percentage of items on the exam that should be allocated to each of the major domains.

In the second activity, sample items were used to illustrate how each standard might be measured as well as to illustrate how standard-specific issues might be addressed. These sample items—which were either available in the public domain or developed by AIR and Insite staff—were aligned with the South Carolina curriculum standards prior to the CRC meetings. The sample items were organized by standard, and previously identified issues were raised with the committees during the discussion of each standard in this activity.

The following is a summary of the types of standard-specific issues that AIR and Insite staff members presented to the CRCs during this activity:

- The decision had been made by the SDE that EOCEP examinations would use a multiple-choice format due to the necessity of reporting student scores before the end of the semester. The contractor reviewed all standards for measurability using multiple-choice items. In most cases, the standards were judged to be measurable by multiple-choice items; however, some standards were judged to be only partially measurable and others not to be measurable at all through multiple-choice items.

- Some standards seemed to cover the same content. AIR and Insite staff needed assistance from the CRCs to differentiate between these standards for the purpose of student assessment as well as to clarify the terminology used in certain standards.

Additionally, AIR and Insite staff asked CRC members to make two decisions during the standards review activity: how important each standard was for inclusion on an end-of-course test and whether the identified taxonomic levels were appropriately aligned with the particular standards. Once the CRC members had made these judgments, the results were used to examine each standard, and any standard on which the group did not reach consensus was discussed further.

Having been given the opportunity to study the standards in greater depth, for the third activity CRCs had the option of reconsidering their initial item allocations. Once all members had made their final judgments, the median item allocation percentage for each domain was computed and presented to the committee. The CRCs agreed to use these percentages as the basis for making subdomain allocations. Applying the committee-determined percentages to the proposed 60-item

field tests, members allocated the number of items for each domain and subdomain. These data were compiled following the CRC meetings and were used to guide initial item writing.

At the CRC meetings, committee members were given sample item review sets to use to evaluate the suitability of the item types for use on EOCEP examinations. The Physical Science committee also was asked to review the proposed periodic table of elements and an equation reference sheet to be used by students during the test.

These preliminary test and item specifications for each subject were comprised of the following:

- SDE decisions on the general issues,
- measurement guidelines reflecting the full discussions of the standards, sample item-review sets reviewed by the CRCs and the CRC subsets, and
- results of the item allocation activity.

1.4 DEVELOPMENT OF FIELD-TEST ITEM POOL

AIR staff members used the measurement guidelines and item specifications developed at the October CRC meetings to train qualified item writers, each of whom had prior item-writing experience. AIR assessment specialists worked with the item writers to explain the purpose of the assessment, review measurement practices in item writing, provide guidance in item writing according to the standards identified as appropriate for each EOCEP exam, and interpret the curriculum standards as necessary.

Sample item types that had been approved by the CRCs served as models for the writers in developing items for the EOCEP tests. To ensure that the items tapped a range of difficulty and taxonomic levels as required by the SDE, the AIR defined the taxonomic levels. A matrix that combined standards and eligible taxonomic levels was then constructed and used to guide item writing. The taxonomic levels were matched to standards based upon the standards themselves and the problem-solving results from the first CRC meeting.

The item writers, who had been trained in item-review criteria, used a checklist created out of those criteria as their guide throughout the item development and review process. The review comprised drafting the test items, receiving feedback from the AIR and Insite assessment specialists, revising the items, and submitting their final drafts to the SDE assessments specialists.

1.5 EXTERNAL ITEM REVIEW

Once reviewed, edited, and approved by the AIR, test items were submitted to the SDE for initial review. AIR staff discussed the SDE's proposed revisions to the items, as necessary, and revised them for the January 2003 CRC and SRC meetings.

A critical step in establishing the validity of the items was the South Carolina educator review of the item pools for content validity and potential bias issues. Once the field-test item pools had been reviewed by SDE Office of Assessment staff members, the items were reviewed by the

CRCs and the SRC. These committees were convened in January 2003 for each of the content areas.

In preparation for the January item-review meetings, the SDE expanded the CRCs and the SRC to include a larger number of district instructional personnel. After a general introductory session, each of the CRC committees was divided into subgroups for the item review, which they conducted during a two-day period each for the biology, English, and physical science EOCEP tests and during a one-day period for the test in algebra. Each committee member was given a secure notebook containing a set of items to be reviewed. In each review group, a lead was identified to direct the review and discussion. The AIR assessment content specialists reviewed the standards and the measurement guidelines from the October CRC meetings with the committees prior to beginning the item reviews. Participants worked out the items, applied item-review criteria, and voted individually to accept, revise, or reject each item. Once all votes were registered, group leaders led discussions on those items for which there was no group consensus and recorded the group's recommendations.

Finally, the CRCs reviewed prototypes of a test form composed of 60 items. The importance ratings and item allocations from the CRCs' October meetings had been used to develop the prototypes, and committee members were asked to review the item allocations and to make any revisions they thought necessary. Once the item allocation results were compiled, the CRCs discussed the results and then, using his or her CRC's consensus domain percentages, each member allocated items to the subdomain level.

The measurement guidelines were updated on the basis of discussion by the CRC members. The committees supported reducing the operational test length from 60 to 55 items. The CRC judged 55 items to be a sufficient number for adequate measurement of the standards at 0.90 reliability. This reduction had the added benefit of decreasing administration time.

Following the CRC meetings, the SRC met to examine all field-test items for potential bias, including language that might disadvantage a group, offend members of a particular group, or present obstacles to a group due to factors unrelated to content and processes specified in the standards. SRC members reviewed items using copies of the same item-review materials used in the CRC meetings, as well as the guidelines for bias, sensitivity, and language simplification developed by the AIR. The AIR bias and sensitivity review leader outlined the purpose of the meeting; discussed the guidelines for bias, sensitivity, and language simplification; and worked through a few of the items aloud with the group before asking members to review a set of items on their own. Members were reminded to concentrate on language rather than content.

After members had completed reviewing the first set of items, the group discussed any items that were identified as potentially problematic. The review leader used the documentation notebooks from the CRC meetings to inform the committee of any items that had been revised or deleted by the CRCs. SRC members reached consensus on item revisions (e.g., change of context, simplification of sentence structure or language for clarity), and the review leader recorded comments and recommendations throughout the meeting.

Following the CRC meetings, the measurement guidelines for English 1, Biology 1/Applied Biology 2, and Physical Science were updated, and blueprints were developed from the item allocation activities. The blueprints used the importance weightings compiled from the CRC

meetings as well as subsequent judgments made by the CRCs regarding item allocations to domains and standards. The CRCs' domain allocations were adjusted to reflect a total operational test length of 55 items and were used in constructing the test blueprint.

1.6 DEVELOPMENT OF FIELD-TEST FORMS

At the conclusion of the item-review process, field-test forms were constructed across the new subjects from the pool of items that were approved during the internal and external review processes. These items, which had been approved by the SDE, were then certified to match the specified content standards.

There were no item statistics to guide the partitioning of items across forms. Rather, the AIR content specialists endeavored to construct forms that mirrored the proposed operational test blueprints.

TABLE 1.1

English 1 Test Blueprint

Strand/Topic	Approx. Number per Form	Approx. Percent per Form
Reading Process and Comprehension	13	24%
Analysis of Texts	14–16	27%
Word Study and Analysis	7–10	16%
Research	2–5	6%
Writing	10–12	20%
Communication	5–7	11%
Total	**55**	

TABLE 1.2

Biology 1/Applied Biology 2 Test Blueprint

Domain	Approx. Number per Form	Approx. Percent per Form
I. Inquiry		
A. Identify Questions	2	4%
B. Design and Conduct Investigations	9	16%
C. Math in Science	0–1	2%
D. Scientific Explanations	0–1	2%
E. Alternative Explanations	0–1	2%
F. Scientific Inquiry	0–1	2%

II. Biology		
A. The Cell	9	16%
B. Heredity	7	13%
C. Biological Evolution	5	9%
D. Interdependence of Organisms	8	15%
E. Matter, Energy and Organization	7	13%
F. Behavior and Regulation	4	7%
Total	**55**	

TABLE 1.3

Physical Science Test Blueprint

Domain	Approx. Number per Form	Approx. Percent per Form
I. Inquiry		
A. Identify Questions	2	4%
B. Design and Conduct Investigations	7	13%
C. Math in Science	3	5%
D. Scientific Explanations	0–1	2%
E. Alternative Explanations	0–1	2%
F. Scientific Inquiry	0–1	2%
II. Chemistry		
A. Structure of Atoms	3	5%
B. Structure and Properties of Matter	11	20%
C. Chemical Reactions	6	11%
III. Physics		
A. Motions and Forces	12	22%
B. Conservation of Energy and the Increase in Disorder	4	7%
C. Interactions of Energy and Matter	4	7%
Total	**55**	

Each field-test form contained 60 multiple-choice items, with a set of anchor items embedded in each form to assure form equivalency and appropriate representation of the standards. These anchor items were placed in the same position on each form in a given subject area. During field-test form construction, consideration was given to content difficulty, match to the blueprint, overlap, and readability load across forms. Once the field-test forms had been constructed, they were sent to the SDE for review and approval. Revisions requested by the SDE were made as agreed upon after review and discussion. The AIR created the final field-test forms for review

and approval by the SDE. A total of 1,954 items were field-tested in the spring 2003 field-test administration.

The spring 2003 field test did not result in a sufficient number of items for pre-equated operational-test forms to be produced for each subject. Additional items were therefore developed and field-tested in English 1, Biology 1/Applied Biology 2, and Physical Science. Implementation tests were administered in fall 2003 and spring 2004.

For the English 1 EOCEP test, the item bank from the spring 2003 field-test administration was supplemented with a total of 42 new items. Of the 44 field-test items, 2 were repeated in two forms, resulting in 42 unique field-test items in the fall 2003 administration and 120 new items in the spring 2004 administration. Thus, for the fall 2003 administration, there were multiple operational forms, all having the same 50 operational items and each having 11 unique embedded field-test items. Similarly, for the spring 2004 administration, there were multiple English 1 operational forms, each with 10 unique field-test items.

For the Biology 1/Applied Biology 2 test, 5 field-test items were added at the end of the operational items in the fall 2003 administration. Of these 5 field-test items, 4 items were refield-tested in the fall 2003 administration because the SDE and the AIR suspected that the inadequate item statistics were due to the lack of student motivation with regard to the field test and that the item statistics might improve in the operational-test setting. However, the item statistics were still inadequate (i.e., low point-biserials) in the fall 2003 administration, and these items were not included in any operational forms.

For the Physical Science test, 5 field-test items were added at the end of the operational items in the fall 2003 and the spring 2004 administrations, creating a total of 10 field-test items. In a situation similar to that which occurred with the biology items, 2 of the 5 physical science items field-tested in fall 2003 were refield-tested in fall 2003. Other field-test items were either new or slightly revised versions of previously field-tested items to fill in some gaps in assessing the content of the standards.

For all subjects, Rasch-ability-score-to-scale-score conversion tables were produced prior to each test administration on the basis of the item parameters in the pre-equated item pool. This technical report summarizes the results of statistical and psychometric analyses performed on the 2003–04 operational data for EOCEP tests in algebra, biology, English, and physical science.

In this report, all data are based on the students in the regular schools and in adult education programs only. Data on students in district-approved home schools have been excluded. (Across the three 2003–04 test administrations, home school students numbered twenty-six in Algebra 1/Mathematics for the Technologies 2, six in English 1, four in Biology 1/Applied Biology 2, and two in Physical Science.)

Chapter 2

STUDENT DEMOGRAPHICS

2.1 STUDENT PARTICIPATION

Operational tests for EOCEP Algebra 1/Mathematics for the Technologies 2 and implementation tests for English 1, Biology 1/Applied Biology 2, and Physical Science were administered in 2003–04. All schools administered these tests to the students who completed courses for Algebra 1, Mathematics for the Technologies 2, Biology 1, Applied Biology 2, Physical Science, or English 1 for credit toward a high school diploma.

Demographic data were collected for each student. These data included the categories of gender, race/ethnicity, grade, English language fluency (LEP, limited English proficiency), lunch program participation, disability status, and migrant status. Table 2.1 presents the combined student participation in the three EOCEP administrations (fall 2003, spring 2004, and summer 2004) by the demographic variables.

TABLE 2.1

Summary of Student Demographics in the Sample

Demographics	Algebra 1/ Math Tech 2		English 1		Biology 1/ Applied Biology 2		Physical Science	
	N	%	N	%	N	%	N	%
Overall	58,017		56,010		37,233		45,817	
Gender								
Female	29,111	50.2	27,582	49.2	19,674	52.8	22,967	50.1
Male	28,602	49.3	27,931	49.9	17,253	46.3	22,552	49.2
Unknown	304	0.5	497	0.9	306	0.8	298	0.7
Grade								
7	1,285	2.2	1	0.0	1	0.0	3	0.0
8	10,586	18.2	5,548	9.9	19	0.1	308	0.7
9	26,272	45.3	49,728	88.8	9,262	24.9	38,877	84.9
10	17,419	30.0	463	0.8	22,740	61.1	5,453	11.9
11	1,503	2.6	81	0.1	3,659	9.8	751	1.6
12	729	1.3	22	0.0	1,440	3.9	277	0.6
Adult education	156	0.3	108	0.2	77	0.2	96	0.2
Other	67	0.1	59	0.1	35	0.1	52	0.1
Ethnicity								
White	32,168	55.4	30,885	55.1	21,870	58.7	25,411	55.5
African American	23,026	39.7	22,031	39.3	13,601	36.5	18,302	39.9
Hispanic	1,362	2.3	1,428	2.5	712	1.9	990	2.2
Asian/Hawaiian-Pac. Islander	590	1.0	602	1.1	461	1.2	438	1.0
American Indian	108	0.2	104	0.2	59	0.2	80	0.2
Other	463	0.8	434	0.8	226	0.6	288	0.6
Unknown	300	0.5	526	0.9	304	0.8	308	0.7
Language								
English speaker	53,723	92.6	52,267	93.3	35,595	95.6	43,855	95.7
Exited	317	0.5	268	0.5	201	0.5	184	0.4
LEP mainstream	108	0.2	91	0.2	40	0.1	71	0.2

TABLE 2.1

Summary of Student Demographics in the Sample

Demographics	Algebra 1/ Math Tech 2		English 1		Biology 1/ Applied Biology 2		Physical Science	
	N	%	N	%	N	%	N	%
Full LEP	402	0.7	535	1.0	197	0.5	305	0.7
Waiver	12	0.0	26	0.0	18	0.0	23	0.1
Other	3,455	6.0	2,823	5.0	1,182	3.2	1,379	3.0
Lunch								
Free meals	18,139	31.3	19,206	34.3	10,332	27.7	15,723	34.3
Reduced-price meals	3,770	6.5	3,810	6.8	2,154	5.8	3,159	6.9
No free/reduced-price meals	33,077	57.0	30,634	54.7	23,721	63.7	25,888	56.5
Unknown	3,031	5.2	2,360	4.2	1,026	2.8	1,047	2.3
IEP								
Yes	4,486	7.7	5,236	9.3	1,845	5.0	3,921	8.6
No	3,329	5.7	3,411	6.1	1,529	4.1	2,619	5.7
Unknown	50,202	86.5	47,363	84.6	33,859	90.9	39,277	85.7
Migrant								
Yes	52	0.1	73	0.1	31	0.1	49	0.1
No	9,669	16.7	10,135	18.1	6,360	17.1	8,475	18.5
Unknown	48,296	83.2	45,802	81.8	30,842	82.8	37,293	81.4
Gifted/talented								
Academic	974	1.7	907	1.6	444	1.2	529	1.2
Artistic	46	0.1	68	0.1	43	0.1	48	0.1
Both	86	0.1	43	0.1	58	0.2	33	0.1
No	6,750	11.6	7,322	13.1	4,413	11.9	6,295	13.7
Unknown	50,161	86.5	47,670	85.1	32,275	86.7	38,912	84.9
504 plan								
Yes	401	0.7	409	0.7	272	0.7	317	0.7
No	53,876	92.9	52,539	93.8	35,196	94.5	43,544	95.0
Unknown	3,740	6.4	3,062	5.5	1,765	4.7	1,956	4.3
Alternative school								
Yes	912	1.6	1,059	1.9	484	1.3	748	1.6
No	52,688	90.8	51,396	91.8	34,478	92.6	42,572	92.9
Unknown	4,417	7.6	3,555	6.3	2,271	6.1	2,497	5.4

2.2 ACCOMMODATIONS

Supplemental information regarding the administration of the EOCEP to students with disabilities is contained in the EOCEP test administration manuals (SDE 2003a, 2004a, and 2004c). These manuals provide guidelines for individualized education program (IEP) teams in making decisions about testing students with disabilities and gives specific information regarding testing accommodations and modifications, test forms and materials, and test administration procedures.

A student with a documented disability is one who has been evaluated and found to meet the eligibility criteria for enrollment in special education as defined by the 1997 amendments to the Individuals with Disabilities Education Act and by State Board of Education Regulation 43-243.1, or one who has a disability covered under Section 504 of the Rehabilitation Act of 1973. The IEP or 504 plan team determines how a student with disabilities participates in the EOCEP assessments. Decisions about accommodations and modifications must be made on an individual student basis and not on the basis of the category of disability. Table 2.2 presents the percentages of accommodations used in the 2003–04 testing.

TABLE 2.2

Accommodations Used in 2003–04 EOCEP Testing

Accommodations	Algebra 1/ Math Tech 2	English 1	Biology 1/ Applied Biology 2	Physical Science
		Regular Form		
	(N = 56,861)	**(N = 54,788)**	**(N = 36,853)**	**(N = 44,923)**
Setting	1.0	0.9	0.6	0.7
Timing	0.1	0.2	0.1	0.0
Scheduling	0.1	0.1	0.0	0.0
Response options	0.0	0.0	0.0	0.0
Presentation	0.0	0.0	0.0	0.0
Other	0.0	0.0	0.0	0.0
		Customized Form		
	(N = 1,156)	**(N = 1,222)**	**(N = 380)**	**(N = 894)**
Setting	61.1	66.6	72.1	68.0
Timing	12.9	10.9	14.5	9.6
Scheduling	4.5	4.3	8.2	5.8
Response options	1.9	1.2	1.6	0.9
Presentation	38.3	30.2	38.2	36.0
Other	0.8	0.4	1.8	1.0

Total responses in each column may exceed 100 percent because some students received accommodations in more than one category.

2.3 TEST ADMINISTRATION TIME

In addition to providing their demographic information, students were asked to record on their answer documents the exact times that they started and finished the test. These answer documents were scanned, and the total elapsed time was calculated for each student.

Across all subjects, 7 to 14 percent of the students who were administered the regular forms and 5 to 16 percent of the students who were administered the customized forms either left one or both times blank or recorded invalid values. Consequently, it was not possible to calculate a total testing time for these students. The percentage of students who finished the test within two hours ranged from 83 to 92 percent for the regular forms and from 79 to 92 percent for the customized forms across subjects, as tables 2.3 and 2.4 reflect.

TABLE 2.3
Time Taken in 2003–04 EOCEP Testing with Regular Forms

Time Taken	Algebra 1/Math Tech 2			English 1		
	Fall 2003 (N = 9,770)	Spring 2004 (N = 45,222)	Summer 2004 (N = 1,869)	Fall 2003 (N = 8,823)	Spring 2004 (N = 44,788)	Summer 2004 (N = 1,177)
15 min	0.6	0.4	0.4	0.4	0.4	1.1
30 min	3.1	2.2	3.0	3.2	1.5	7.4
45 min	14.3	11.5	10.3	16.5	6.8	22.7
1 hr	28.9	25.8	22.9	30.2	20.5	27.4
1 hr 15 min	19.9	22.7	21.2	18.8	22.3	17.4
1 hr 30 min	9.9	14.2	17.0	9.2	15.5	8.2
1 hr 45 min	5.2	7.1	9.0	4.6	9.9	3.6
2 hr	2.4	3.7	4.7	2.3	6.3	2.6
2 hr 15 min	1.0	1.7	2.2	1.0	3.2	0.8
2 hr 30 min	0.4	0.8	1.2	0.4	1.6	0.4
2 hr 45 min	0.2	0.4	0.6	0.2	0.7	—
3 hr or more	0.4	0.5	0.4	0.2	0.8	—
Invalid*	13.7	9.0	7.1	13.1	10.6	8.5
Time Taken	**Biology 1/Applied Biology 2**			**Physical Science**		
	Fall 2003 (N = 10,461)	**Spring 2004 (N = 26,278)**	**Summer 2004 (N = 114)**	**Fall 2003 (N = 12,852)**	**Spring 2004 (N = 31,965)**	**Summer 2004 (N = 106)**
15 min	1.1	3.3	1.8	1.0	1.2	—
30 min	17.8	33.0	27.2	7.5	8.4	16.0
45 min	36.8	36.6	32.5	25.9	26.8	31.1
1 hr	21.7	13.2	15.8	30.4	29.1	28.3
1 hr 15 min	6.3	3.2	6.1	12.9	14.5	9.4
1 hr 30 min	2.0	1.4	3.5	4.6	5.9	0.9
1 hr 45 min	1.0	0.7	—	2.3	2.4	0.9
2 hr	0.4	0.2	—	1.2	1.3	—
2 hr 15 min	0.1	0.1	0.9	0.6	0.5	—
2 hr 30 min	0.1	0.0	—	0.2	0.2	—
2 hr 45 min	0.0	0.0	—	0.1	0.1	0.9
3 hr or more	0.0	0.0	—	0.1	0.1	—
Invalid*	12.8	8.2	12.3	13.3	9.4	12.3

* includes responses with no mark or double marks on start and stop time fields so that it was not possible to compute the difference between start and stop times

TABLE 2.4

Time Taken in 2003–04 EOCEP Testing with Customized Forms

Time Taken	Algebra 1/Math Tech 2			English 1		
	Fall 2003 (N = 290)	Spring 2004 (N = 829)	Summer 2004 (N = 37)	Fall 2003 (N = 202)	Spring 2004 (N = 976)	Summer 2004 (N = 44)
15 min	—	1.1	—	—	0.8	—
30 min	2.1	3.1	2.7	0.5	2.4	—
45 min	5.9	7.6	10.8	1.5	3.7	—
1 hr	14.8	19.2	10.8	13.9	16.0	2.3
1 hr 15 min	14.5	18.9	21.6	25.7	33.3	36.4
1 hr 30 min	24.8	13.3	21.6	28.7	18.1	27.3
1 hr 45 min	14.1	7.0	2.7	9.4	11.8	18.2
2 hr	7.2	8.4	13.5	7.4	1.1	11.4
2 hr 15 min	3.1	2.7	5.4	1.5	1.4	—
2 hr 30 min	3.8	2.3	—	0.5	0.5	—
2 hr 45 min	0.7	1.1	—	—	0.6	—
3 hr or more	0.7	1.2	—	—	0.5	—
Invalid*	8.3	14.1	10.8	10.9	9.7	4.5
Time Taken	**Biology 1/Applied Biology 2**			**Physical Science**		
	Fall 2003 (N = 75)	**Spring 2004 (N = 303)**	**Summer 2004 (N = 2)**	**Fall 2003 (N = 284)**	**Spring 2004 (N = 607)**	**Summer 2004 (N = 3)**
15 min	—	2.3	—	0.4	0.8	—
30 min	4.0	4.6	—	1.4	4.4	—
45 min	21.3	28.1	—	12.0	17.5	33.3
1 hr	29.3	30.0	100	22.2	31.0	33.3
1 hr 15 min	20.0	12.9	—	19.7	20.6	—
1 hr 30 min	4.0	4.3	—	14.1	6.1	—
1 hr 45 min	2.7	6.9	—	10.6	3.6	—
2 hr	2.7	3.0	—	2.1	4.4	—
2 hr 15 min	1.3	0.3	—	1.1	1.8	—
2 hr 30 min	—	0.3	—	—	—	—
2 hr 45 min	1.3	0.3	—	0.4	—	—
3 hr or more	—	0.3	—	0.7	0.2	—
Invalid*	13.3	6.6	—	15.5	9.6	33.3

* includes responses with no mark or double marks on start and stop time fields so that it was not possible to compute the difference between start and stop times

2.4 STUDENT QUESTIONNAIRE

After the administration of the EOCEP test in each subject, students were instructed to complete a questionnaire that addressed such topics as the difficulty of the test, the nature of the instruction they had received in the particular course, their use of calculators in the particular course (algebra and physical science only), and the amount of time they had spent engaged in lab activities in the particular course (biology and physical science only).

Chapter 3

TEST ADMINISTRATION

3.1 TEST ADMINISTRATION WINDOW

The fall 2003 testing window ran from December 8, 2003, through January 21, 2004. School districts were given the option of administering all EOCEP tests within a single five-day period or within two five-day periods. Districts were instructed to administer makeup tests following their regular testing period. For all three EOCEP administrations—fall 2003, spring 2004, and summer 2004—district test coordinators (DTCs) were responsible for providing the testing schedule to all school test coordinators (STCs) in their particular districts.

The spring 2004 state testing window ran from May 3, 2004, through June 8, 2004; the summer 2004 testing window ran from June 21, 2004, through August 3, 2004. Districts were again allowed to choose either one five-day period or two five-day testing periods for EOCEP testing in all subjects. Whether a district opted for a single five-day testing period or for two five-day testing periods, it was required by the state to administer the test for each separate EOCEP subject within a single five-day period.

For students in Algebra 1 or Mathematics for the Technologies 2 who missed the originally scheduled EOCEP test due to a death in the family, illness, or another situation deemed valid by the state, school districts were required to have a five-day makeup period the week immediately following the original test administration. The remaining subjects were at the implementation stage and were not required to have a separate makeup period; any makeup tests for these subjects had to be administered during the original five-day testing period. For these subjects, it was recommended that a single make-up test be given per day, but two may have been given per day if necessary.

3.2 TIMING OF THE TEST

The EOCEP tests were not timed; however, each session had to be administered during a single day (unless a student's IEP or 504 plan specifically stated that he or she needed to have the test administered over several days). Districts and schools were instructed that students should be given as much uninterrupted time as necessary to take the test to ensure an accurate assessment. For planning purposes, districts and schools were informed that approximately 80 to 90 percent of students from the May 2003 EOCEP administration completed each test within two hours.

3.3 ADMINISTRATION MANUALS

Working with the SDE, AIR staff drafted the administration manuals for the test. SDE staff reviewed and revised the manuals, and the AIR finalized and printed them. The EOCEP district test coordinator supplements (SDE 2003b, 2004b, and 2004d) and the EOCEP test administration manuals (TAMs) were produced for each administration of the EOCEP tests (fall 2003, spring 2004, and summer 2004). The DTC supplements included only the information that DTCs needed for the administration of the EOCEP tests. The TAMs contained the information that

STCs, test administrators (TAs), and monitors needed to administer the tests to students in their schools.

The TAMs and the supplements included logistical and administration procedures as well as the directions (scripts) for administering the tests. The DTCs, STCs, and TAs were encouraged to use a form provided in the manuals to offer comments and suggestions on the procedures and both the TAM and the supplement. The comments were compiled in a spreadsheet and sent to the SDE to review and to use as the basis for potential changes in test procedures. The TAM also included a testing irregularity form that test administrators were instructed to use to report any problems or deviations from established testing procedures.

For the spring and summer 2004 administrations, the TAM included additional graphics that had been suggested on the comment forms from previous administrations. In addition, appendix C in the 2004 TAM was revised to include a more detailed description of materials available, as well as additional graphics for completing student demographic information and returning scorable and nonscorable test materials. Tables showing the types of customized materials available for students who require such special testing formats were also provided.

3.4 CUSTOMIZED MATERIALS

Customized formats of the EOCEP test were available for Algebra 1/Mathematics for the Technologies 2, Biology 1/Applied Biology 2, Physical Science, and English 1:

- Loose-leaf test booklets—printed single-sided, one item to a page, and bound in three-ring binders—allowed individuals to remove the pages, if necessary, during testing.
- Large-print booklets were produced for students who have difficulty reading text in a standard-size font. The large-print version used an 18-point sans serif font and was issued as a 9 x 12-inch spiral-bound booklet.
- Braille booklets were produced for students who typically read classroom materials in braille. The braille version was issued as spiral-bound booklet containing 11½ x 11-inch interpoint braille pages.
- A regular print Form C test booklet was provided in test packets for students or TAs to use with customized formats such as the oral script, braille, large-print, loose-leaf, and sign language versions. These booklets were saddle-stitched and printed in a 12-point font, just as the regular, noncustomized test booklets were.
- For students whose IEP or 504 plan requires the oral administration of tests, oral administration scripts gave specific directions to TAs regarding the appropriate way to read test questions, passages, and response options.
- Sign language videotapes—produced for English 1, Biology 1/Applied Biology 2, and Physical Science—included the signed test directions, questions, and response options. The videotapes were produced in two languages: American Sign Language and Pidgin Signed English.
- For the Algebra 1/Mathematics for the Technologies 2 test, a sign language script was provided for American Sign Language, Pidgin Signed English, and Signed Exact English test-takers.

3.5 PRETEST WORKSHOPS AND TRAINING

Pretest workshops were scheduled prior to each test administration window to train the DTCs and some STCs. DTCs were allowed to bring up to three additional representatives to the workshop. SDE and AIR staff conducted the training. The fall 2003–winter 2004 workshops were held on October 2–3, 2003. The spring 2004 workshops were held on March 29–31, 2004. The summer 2004 workshops were held on June 9, 2004. The workshop sessions included overviews of the EOCEP as well as detailed discussions by AIR staff on the details of materials handling, check-in and return procedures, testing windows, test timing, makeup procedures, calculator use policy, security requirements, and accommodations and modifications for students with disabilities.

The fall 2003 administration manuals were distributed at the workshops. All test coordinators received a copy of the TAM; the DTCs also received a copy of the supplement. For the spring 2004 administration, DTC supplements and TAMs were mailed to the DTCs two weeks prior to the pretest workshops to allow time for them to distribute the manuals to attendees and read the manuals before the workshops. For the spring 2004 and summer 2004 workshops, attendees also received printed copies of PowerPoint presentations given by AIR and PEM staff.

These PowerPoint presentations were also posted to the SDE Web site to allow the DTCs to use them in their training sessions with the STCs, who would themselves be training the TAs and monitors in their respective schools at least one week before test day. Monitors were strongly recommended for any classroom with twenty-five or more students test-takers.

3.6 MATERIALS SHIPPING AND RETURN

For all three administrations, test materials were shipped to district offices approximately two weeks before testing—in time for the DTCs to distribute school materials at least one week before the schools' test dates. Each school's shipment was boxed individually and labeled with the total number of boxes shipped to that school.

The district office was also sent a shipment of noncustomized overage materials, which were to be used by the DTCs to complete any additional materials requests from the STCs. Materials in customized formats were sent only to the schools and only in the quantities ordered.

TAs were instructed to return their test materials to the STCs immediately after the test administration. The STCs then redistributed test materials to the TAs who needed them in order to administer makeup tests. Those TAs were instructed to return the makeup test materials to their STC immediately after the makeup session. DTCs were to arrange for the pickup of all scorable materials for return to PEM within three days after testing.

Because the 2003–04 Algebra 1/Mathematics for the Technologies 2 test scores were required to be reported back to the schools quickly for calculating final course grades, a rapid scoring and reporting process was implemented for all three administrations. Each school district could return the scorable materials to PEM, in as many as five separate shipments, as they arrived from the schools. Nonscorable materials were to be returned in one shipment within three days of the completion of makeup tests. For all three administrations, step-by-step instructions for returning scorable and nonscorable materials were included in the district materials. These directions listed the toll-free phone numbers of the trucking companies that the DTCs were instructed to call to schedule return materials pickups.

Chapter 4

TECHNICAL CHARACTERISTICS OF ITEMS

This chapter reports the results of item analyses based on classical test theory (CTT) using a proprietary program designed by the AIR. Item difficulty (*p*) is the proportion (or percentage) of examinees correctly answering a dichotomously scored item.

Item discrimination is defined as a correlation between the item score and the total score. For the discrimination index, point-biserial correlations were produced. In computing the point-biserial correlation, the AIR corrected for spuriousness. In the recoding of missing data for item analysis, all omitted and not-reached items were recoded as incorrect, with a zero score. After discussions between the SDE and the AIR, it was decided to exclude from the CTT item analyses and item calibrations those students who had used customized test materials.

4.1 ITEM NONRESPONSE RATES

Although the EOCEP tests were not timed, students were required to finish each test during one school day, unless they had an IEP that allowed for accommodations in administration. Districts and schools were instructed that, if they had space and staff available, students should be given as much uninterrupted time as necessary to take the test to ensure an accurate assessment.

The item nonresponse rates indicate the percentage of students who did not reach a particular item and all items thereafter. The item omit rates indicate the percentage of students who did not respond to that particular item but did respond to a later item. In English 1, the not-reached percentage for the last item ranged from 1.30 percent to 3.26 percent across forms. In Biology 1/Applied Biology 2, Physical Science, and Algebra 1/Mathematics for the Technologies 2, the not-reached percentage was less than 1 percent for the last item in all forms, with the exception of the fall 2003 Physical Science administration (1.49 percent). The omit rates were quite low in all subjects, ranging from 0 percent to 2.63 percent. These data indicate that students in all subjects were given ample time to complete the test.

4.2 CLASSICAL ITEM STATISTICS

Table 4.1 provides a summary of item *p*-values and item discrimination values for operational, implementation, and field-test items for the fall 2003 and spring 2004 administrations. The item statistics for the summer administrations were not calculated because the same form (Form 10) was administered in fall 2003 and summer 2004 for the English 1, Biology 1/Applied Biology 2, and Physical Science tests. For Algebra 1/Mathematics for the Technologies 2, the summer form (Form 10) was administered in fall 2002. The point-biserial correlations were corrected for spuriousness.

TABLE 4.1

Summary of Classical Item Statistics

Administration	Number of Items	*p*-value	Adjusted Point-Biserial Correlation
Algebra 1/Math Tech 2			
Fall 2003 (Form 30)	50	0.51	0.28
Spring 2004 (Form 40)	50	0.60	0.35
English 1			
Fall 2003 (Form 10)*	50	0.62	0.37
Spring 2004 (Form 20)	55	0.65	0.36
Fall 2003 field-test items	42	0.52	0.33
Spring 2004 field-test items	120	0.53	0.38
Biology 1/Applied Biology 2			
Fall 2003 (Form 10)	55	0.53	0.29
Spring 2004 (Form 40)	55	0.51	0.30
Fall 2003 field-test items	5	0.36	0.10
Physical Science			
Fall 2003 (Form 10)	55	0.47	0.28
Spring 2004 (Form 20)	55	0.46	0.26
Fall 2003 field-test items	5	0.33	0.15
Spring 2004 field-test items	5	0.42	0.20

* The fall 2004 English 1 test form was considered as a preliminary Form 10 because that form included only 50 items. For the final Form 10, 5 items were added, bringing the total for the future operational administrations to 55 items.

Chapter 5

ITEM CALIBRATION AND SCALING

5.1 METHODOLOGY AND SOFTWARE

The one-parameter Rasch dichotomous model (Rasch 1960; Wright and Stone 1979) was used to calibrate multiple-choice items. WINSTEPS software (see Linacre and Wright 2003) was used in the item calibration. The program employs joint maximum likelihood estimation (JMLE), an approach that estimates the item and person parameters simultaneously.

5.2 PRE-EQUATING

The AIR conducted field tests with a sufficient number of items in spring 2003 to create a precalibrated item pool and to construct the pre-equated operational-test forms from the precalibrated item pool for English 1, Biology 1/Applied Biology 2, and Physical Science. As explained in section 5.3, below, the precalibrated item pool for English 1 and Physical Science were supplemented in the fall 2003 and spring 2004 administrations. For all subjects, the Rasch-ability-score-to-scale-score conversion tables were produced prior to each test administration based on the item parameters in the pre-equated item pool.

For Algebra 1/Mathematics for the Technologies 2 EOCEP testing, the operational-test forms were constructed from the precalibrated item pool based on the spring 2002 field test and embedded field-test items in the fall 2002 and spring 2003 administrations; therefore, the Rasch-ability-score-to-scale-score conversion tables for the fall 2003, spring 2004, and summer 2004 operational-test forms were created before the test administration.

5.3 ITEM CALIBRATION

For English 1 EOCEP testing, the item bank from the spring 2003 field-test administration was supplemented with a total of 162 new items that were embedded among the operational items or added at the end of the operational items. In the fall 2003 administration, four English operational forms were administered. The test's 61 multiple-choice items included 50 operational items and 11 embedded field-test items, resulting in a total of 42 unique field-test items for future use. In the spring 2004 administration, twelve operational forms were administered. Ten field-test items were added after the operational items in each form, resulting in a total of 120 field-test items. For each administration, all English 1 test forms were spiraled within a classroom.

In the fall 2003 and spring 2004 English 1 test administrations, in placing the field-test items on the item bank scale, the operational item parameters were anchored at the bank difficulty values; the field-test item difficulties were mapped onto the bank metric in the concurrent calibration.

For Biology 1/Applied Biology 2 EOCEP testing, 5 field-test items were added at the end of the operational items in the fall 2003 administration. Of the 5 field-test items, 4 were refield-tested in the fall 2003 administration because the SDE and the AIR suspected that the inadequate item statistics were due to the lack of student motivation with regard to the field test and that the item

statistics might improve in the operational-test setting. However, the resultant item statistics were still inadequate (i.e., low point-biserials) in the fall 2003 administration.

For Physical Science EOCEP testing, 5 field-test items were added at the end of the operational items in the fall 2003 and the spring 2004 administrations, creating a total of 10 field-test items. In a situation similar to that which occurred with the biology items, 2 of the 5 physical science items field-tested in fall 2003 were refield-tested in fall 2003. Biserial correlations were compared for these items because biserial correlations had been computed for the pre-equated item pool. The biserial correlations improved from 0.02 to 0.17 for one of these items and from 0.07 to 0.32 for the other. Due to the need to fill in some gaps in assessing the content of standards, certain other field-test items either were new or were slightly revised versions of previously field-tested items.

The field-test items in the fall 2003 and spring 2004 Biology 1/Applied Biology 2 and Physical Science test forms were calibrated anchoring on the operational item parameters at the bank difficulty values; the field-test item difficulties were mapped onto the bank metric in the concurrent calibration. For English 1, the calibration included a total of 7,497 students in the fall 2003 administration and 36,299 students in the spring 2004 administration. For Biology 1/Applied Biology 2, the calibration included a total of 9,314 students in the fall 2003 administration and 11,107 students in the spring 2004 administration. For Physical Science, the calibration included a total of 25,759 students in the spring 2004 administration.

5.4 SCALING

The SDE provided the AIR with initial Rasch-ability-score-to-scale-score conversion tables that showed the transformation of the ability score interval for each scale score for each subject area. The AIR then generated these tables specifically for each test form for each subject area on the basis of the pre-equated item pool. The scale scores took into account any differences in the difficulty of the forms due to pre-equating; that is, all items shared a common metric so that the scale scores developed for each form were automatically adjusted for differences in item difficulty. For all subjects, the scale scores ranged from 0 to 100, and the passing scale score cut point was set at 70.

5.5 DEFINITION OF SCOREABILITY

A student was considered “tested” if the student answered at least one question in the answer document. All tested students’ item responses were scored. All omits and not-reached items were recoded as incorrect, with a zero score.

5.6 REPORTING OF ZERO AND PERFECT SCORES

In item response theory (IRT) maximum-likelihood ability estimation methods, zero and perfect scores are assigned the ability of minus and plus infinity. The AIR used the WINSTEPS default setting in estimating finite values for the extreme scores. In other words, a fractional score point value was subtracted from perfect scores, and was added to zero score. The WINSTEPS default value for adjusting the extreme scores for extreme measures is 0.3.

5.7 CUT SCORES FOR PASSING SCORES

After the spring 2003 EOCEP field test, the Buros Institute for Assessment Consultation and Outreach (BIACO) conducted the standard setting workshops for English 1, Biology 1/Applied Biology 2, and Physical Science examinations on July 28–30, 2003. The passing scores recommended by the panelists at the 2003 standard setting workshops were presented to the SDE's technical advisory committee and the SDE. After reviewing the results, the SDE provided the AIR with a table of Rasch-ability-score-to-scale-score transformations for these subjects. The cut scores for the 2003–04 EOCEP test forms are presented in table 5.1.

TABLE 5.1

Cut Scores on the Rasch Ability Scale and the Reporting Scale

Form	Rasch Ability	Scale Score
Algebra 1/Math Tech 2		
Form 30	-0.510	70
Form 40	-0.539	70
Form 10	-0.534	70
English 1		
Form 10	0.482	71
Form 20	0.445	70
Biology 1/Applied Biology 2		
Form 10	-0.043	70
Form 40	-0.048	70
Physical Science		
Form 10	-0.131	70
Form 20	-0.064	71

5.8 PERCENTAGE OF STUDENTS SCORING IN EACH LETTER-GRADE EQUIVALENT

Tables 5.2 through 5.9 report student performance on the Algebra 1/Mathematics for the Technologies 2 operational tests and on the English 1, Biology 1/Applied Biology 2, and Physical Science implementation tests for the three EOCEP test administrations (fall 2003, spring 2004, and summer 2004) combined. The results are summarized separately for regular schools and for adult education programs. The number and percentage of students in each letter-grade equivalent (A, B, C, D, and F) and the mean scale score are reported for the test-takers overall and by demographics. The letter-grade equivalent is the letter grade that corresponds to the student's scale score, which is reported according to its conversion on the South Carolina uniform grading scale.

TABLE 5.2

Algebra 1/Math Tech 2 Operational Test, Regular Schools: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Overall	57,861	79.4	16.6	16.2	25.3	21.4	20.5
Gender							
Female	29,040	79.7	16.2	16.6	26.5	21.7	19.0
Male	28,522	79.3	17.1	15.9	24.2	21.1	21.8
Unknown	299	72.1	5.0	7.0	19.7	25.4	42.8
Grade							
7	1,285	92.9	61.6	19.4	14.0	3.5	1.5
8	10,586	89.9	45.9	24.7	21.0	6.4	2.0
9	26,272	79.3	12.7	18.3	29.4	21.2	18.3
10	17,419	73.0	2.8	8.4	22.8	31.2	34.8
11	1,503	73.6	4.4	9.6	23.2	28.6	34.3
12	729	75.0	5.6	11.5	25.2	30.0	27.6
Other	67	72.1	9.0	10.4	11.9	17.9	50.7
Ethnicity							
White	32,115	82.7	23.4	19.8	26.4	17.8	12.5
African American	22,932	74.9	6.9	11.4	24.0	26.4	31.3
Hispanic	1,362	78.0	13.7	14.0	24.3	24.6	23.4
Asian/Hawaiian-Pac. Islander	588	87.4	41.8	17.7	22.1	10.2	8.2
American Indian	108	77.1	11.1	10.2	27.8	26.9	24.1
Other	458	76.6	10.7	12.4	23.6	25.5	27.7
Unknown	298	72.4	4.7	7.7	19.8	25.5	42.3
Language							
English speaker	53,630	79.7	16.9	16.5	25.6	21.1	19.8
Exited	317	81.0	24.3	12.6	25.2	19.6	18.3
LEP mainstream	108	79.5	15.7	16.7	27.8	19.4	20.4
Full LEP	402	75.0	9.5	8.7	24.6	23.9	33.3
Waiver	12	81.3	16.7	8.3	50.0	16.7	8.3
Other	3,392	76.3	11.0	12.0	21.8	26.0	29.2

TABLE 5.2

Algebra 1/Math Tech 2 Operational Test, Regular Schools:
Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Lunch							
Free meals	18,109	75.4	7.7	11.7	25.1	25.7	29.8
Reduced-price meals	3,767	78.6	12.9	16.2	27.5	23.4	20.0
No free/reduced-price meals	33,026	82.1	22.7	19.2	25.6	18.1	14.4
Unknown	2,959	74.6	6.9	9.6	21.8	29.7	31.9
IEP							
Yes	4,481	70.6	2.9	6.2	15.9	27.3	47.7
No	3,241	77.0	12.8	14.0	21.7	23.5	28.1
Unknown	50,139	80.4	18.0	17.2	26.4	20.8	17.5
Migrant							
Yes	52	77.4	17.3	13.5	11.5	25.0	32.7
No	9,584	77.0	12.0	14.0	23.5	23.0	27.5
Unknown	48,225	79.9	17.5	16.6	25.7	21.1	19.0
Courses taken							
2111 (Alg 1, grade 7 or 8)	8,990	89.9	46.5	24.0	20.9	6.3	2.3
3142 (Math Tech 2)	18,743	72.8	2.8	8.0	22.6	30.8	35.8
4111 (Alg 1)	27,752	80.4	15.7	19.4	29.0	20.1	15.8
Other	2,376	80.2	22.4	14.1	21.3	19.7	22.5
Gifted/talented							
Academic	974	89.1	42.0	25.3	23.8	6.0	3.0
Artistic	46	82.1	17.4	21.7	32.6	17.4	10.9
Both	86	88.7	41.9	24.4	23.3	5.8	4.7
No	6,666	75.0	7.5	11.8	22.8	25.3	32.6
Unknown	50,089	79.8	17.2	16.6	25.7	21.3	19.2
504 plan							
Yes	401	77.4	14.5	13.7	23.9	23.2	24.7
No	53,788	79.7	17.1	16.6	25.6	21.0	19.7
Unknown	3,672	75.4	9.2	11.2	21.2	27.2	31.2
Alternative school							
Yes	858	71.5	4.0	7.8	19.2	23.9	45.1
No	52,658	79.9	17.4	16.7	25.8	20.8	19.2
Unknown	4,345	75.3	8.4	11.1	21.0	28.7	30.7

TABLE 5.3

Algebra 1/Math Tech 2 Operational Test, Adult Education Programs: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Overall	156	69.1	2.6	4.5	14.7	20.5	57.7
Gender							
Female	71	69.8	2.8	2.8	18.3	22.5	53.5
Male	80	68.9	2.5	6.3	12.5	18.8	60.0
Unknown	5	62.4	0.0	0.0	0.0	20.0	80.0
Grade							
9	30	71.9	3.3	6.7	33.3	10.0	46.7
10	23	70.4	4.3	4.4	13.0	26.1	52.2
11	9	66.4	0.0	0.0	22.2	22.2	55.6
12	65	70.3	3.1	6.15	12.3	24.6	53.8
Other	29	63.3	0.0	0.0	0.0	17.2	82.8
Ethnicity							
White	53	73.2	7.5	7.5	20.8	20.8	43.4
African American	94	66.4	0.0	1.1	10.6	21.3	67.0
Asian/Hawaiian-Pac. Islander	2	83.0	0.0	50.0	50.0	0.0	0.0
Other	5	74.0	0.0	20.0	20.0	20.0	40.0
Unknown	2	63.0	0.0	0.0	0.0	0.0	100.0
Language							
English speaker	93	70.6	4.3	5.4	19.4	20.4	50.5
Other	63	66.9	0.0	3.2	7.9	20.6	68.3
Lunch							
Free meals	30	70.9	3.3	3.3	26.7	20.0	46.7
Reduced-price meals	3	66.3	0.0	0.0	0.0	33.3	66.7
No free/reduced-price meals	51	72.0	5.9	7.8	19.6	21.6	45.1
Unknown	72	66.4	0.0	2.8	6.9	19.4	70.8
IEP							
Yes	5	67.6	20.0	0.0	0.0	0.0	80.0
No	88	71.0	3.4	5.7	20.5	23.9	46.6
Unknown	63	66.6	0.0	3.2	7.9	17.5	71.4
Migrant							
No	85	71.3	4.7	5.9	21.2	21.2	47.1
Unknown	71	66.5	0.0	2.8	7.0	19.7	70.4
Courses taken							
3142 (Math Tech 2)	82	68.8	2.4	4.9	17.1	15.9	59.8
4111 (Alg 1)	52	69.6	3.8	3.9	11.5	25.0	55.8
Other	22	69.2	0.0	4.6	13.6	27.3	54.5
Gifted/talented							
No	84	71.4	4.8	6.0	21.4	21.4	46.4
Unknown	72	66.4	0.0	2.8	6.94	19.4	70.8
504 plan							

TABLE 5.3

Algebra 1/Math Tech 2 Operational Test, Adult Education Programs: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
No	88	71.1	4.5	5.7	20.5	21.6	47.7
Unknown	68	66.5	0	2.9	7.35	19.1	70.6
Alternative school							
Yes	54	71.1	3.7	5.6	25.9	18.5	46.3
No	30	71.9	6.7	6.7	13.3	26.7	46.7
Unknown	72	66.4	0.0	2.8	6.9	19.4	70.8

TABLE 5.4

English 1 Implementation Test, Regular Schools: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Overall	55,902	72.3	6.6	12.8	18.6	20.4	41.6
Gender							
Female	27,536	73.7	7.1	13.9	20.2	21.5	37.2
Male	27,869	71.1	6.1	11.8	17.1	19.3	45.7
Unknown	497	67.4	1.6	7.8	14.3	19.7	56.5
Grade							
7	1	77.0	0.0	0.0	100.0	0.0	0.0
8	5,548	84.0	19.8	30.2	29.0	15.4	5.6
9	49,728	71.2	5.2	11.0	17.6	21.0	45.2
10	463	61.7	0.6	2.6	7.8	15.3	73.7
11	81	59.4	1.2	1.2	3.7	12.3	81.5
12	22	70.4	0.0	18.2	18.2	13.6	50.0
Other	59	62.9	1.7	3.4	6.8	18.6	69.5
Ethnicity							
White	30,830	76.2	10.3	18.3	22.6	19.7	29.0
African American	21,985	67.2	1.5	5.4	13.5	21.6	58.0
Hispanic	1,425	66.9	3.4	7.7	14.3	17.3	57.2
Asian/Hawaiian-Pac. Islander	602	77.1	14.5	17.9	19.9	17.6	30.1
American Indian	104	69.7	4.8	9.6	14.4	24.0	47.1
Other	430	70.2	5.6	9.5	15.1	20.0	49.8
Unknown	526	67.2	1.3	7.0	12.9	20.7	58.0
Language							
English speaker	52,162	72.7	6.8	13.2	19.0	20.6	40.4
Exited	267	72.0	7.5	12.4	18.7	15.7	45.7
LEP mainstream	91	65.1	0.0	5.5	14.3	15.4	64.8
Full LEP	535	57.3	0.2	0.7	3.0	8.8	87.3
Waiver	26	69.0	0.0	7.7	26.9	19.2	46.2

TABLE 5.4

English 1 Implementation Test, Regular Schools:
Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Other	2,821	68.4	3.5	8.2	14.3	19.7	54.3
Lunch							
Free meals	19,140	67.2	1.5	5.7	13.4	21.3	58.0
Reduced-price meals	3,806	70.9	3.5	9.6	17.6	24.1	45.2
No free/reduced-price meals	30,603	76.1	10.4	18.1	22.4	19.5	29.6
Unknown	2,353	67.2	2.5	7.2	12.9	18.9	58.4
IEP							
Yes	5,229	61.2	0.5	2.5	5.8	12.3	78.9
No	3,315	70.1	4.6	9.9	16.5	20.5	48.5
Unknown	47,358	73.7	7.4	14.1	20.2	21.3	37.0
Migrant							
Yes	73	64.0	5.5	1.4	17.8	11.0	64.4
No	10,031	69.6	4.1	9.6	15.8	20.5	50.1
Unknown	45,798	72.9	7.1	13.5	19.2	20.4	39.7
Courses taken							
2011 (Eng 1, grade 7 or 8)	4,318	82.3	17.3	27.9	28.4	15.9	10.4
3011 (Eng 1)	50,063	71.6	5.7	11.6	17.8	20.8	44.0
Other	1,521	69.8	5.9	8.9	16.4	18.7	50.1
Gifted/talented							
Academic	907	80.3	17.5	22.4	25.2	15.8	19.1
Artistic	68	78.8	22.1	14.7	17.6	20.6	25.0
Both	43	83.1	16.3	25.6	37.2	16.3	4.7
No	7,220	67.5	2.1	6.7	13.8	20.8	56.6
Unknown	47,664	72.9	7.0	13.5	19.2	20.4	39.8
504 plan							
Yes	408	70.0	4.4	11.3	15.2	16.9	52.2
No	52,438	72.6	6.8	13.1	19.0	20.5	40.6
Unknown	3,056	67.6	2.7	7.5	13.3	19.6	56.9
Alternative school							
Yes	960	64.2	1.8	3.8	10.2	17.5	66.8
No	51,392	72.8	6.9	13.3	19.1	20.5	40.1
Unknown	3,550	67.6	2.5	7.4	13.5	20.0	56.7

TABLE 5.5

English 1 Implementation Test, Adult Education Programs:
Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Overall	108	62.5	0.9	3.7	8.3	13.0	74.1

TABLE 5.5

English 1 Implementation Test, Adult Education Programs: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Gender							
Female	46	64.6	2.2	2.2	8.7	13.0	73.9
Male	62	60.9	0.0	4.8	8.1	12.9	74.2
Grade							
9	94	62.4	1.1	4.3	8.5	11.7	74.5
10	5	61.6	0.0	0.0	0.0	20.0	80.0
11	4	63.8	0.0	0.0	0.0	25.0	75.0
12	3	56.7	0.0	0.0	0.0	0.0	100.0
Other	2	78.0	0.0	0.0	50.0	50.0	0.0
Ethnicity							
White	55	63.4	1.8	3.6	10.9	14.5	69.1
African American	46	61.1	0.0	2.2	6.5	8.7	82.6
Hispanic	3	58.0	0.0	0.0	0.0	33.3	66.7
Other	4	69.5	0.0	25.0	0.0	25.0	50.0
Language							
English speaker	105	62.3	1.0	3.8	8.6	11.4	75.2
Exited	1	72.0	0.0	0.0	0.0	100.0	0.0
Other	2	69.5	0.0	0.0	0.0	50.0	50.0
Lunch							
Free meals	66	60.6	0.0	4.5	3.0	10.6	81.8
Reduced-price meals	4	72.5	0.0	0.0	50.0	0.0	50.0
No free/reduced-price meals	31	65.9	3.2	3.2	16.1	16.1	61.3
Unknown	7	59.4	0.0	0.0	0.0	28.6	71.4
IEP							
Yes	7	57.1	0.0	0.0	14.3	0.0	85.7
No	96	63.0	1.0	4.2	8.3	12.5	74.0
Unknown	5	61.2	0.0	0.0	0.0	40.0	60.0
Migrant							
No	104	62.7	1.0	3.8	8.7	12.5	74.0
Unknown	4	58.3	0.0	0.0	0.0	25.0	75.0
Courses taken							
2011 (Eng 1, grade 7 or 8)	1	73.0	0.0	0.0	0.0	100.0	0.0
3011 (Eng 1)	103	62.5	1.0	3.9	8.7	11.7	74.8
Other	4	60.8	0.0	0.0	0.0	25.0	75.0
Gifted/talented							
No	102	62.5	1.0	3.9	8.8	11.8	74.5
Unknown	6	62.0	0.0	0.0	0.0	33.3	66.7
504 plan							
Yes	1	61.0	0.0	0.0	0.0	0.0	100.0
No	101	62.6	1.0	4.0	8.9	11.9	74.3
Unknown	6	61.7	0.0	0.0	0.0	33.3	66.7
Alternative school							
Yes	99	62.5	1.0	4.0	8.1	12.1	74.7
No	4	65.0	0.0	0.0	25.0	0.0	75.0
Unknown	5	61.2	0.0	0.0	0.0	40.0	60.0

TABLE 5.6

Biology 1/Applied Biology 2 Implementation Test, Regular Schools: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Overall	37,156	70.2	2.7	7.3	18.6	22.4	49.0
Gender							
Female	19,634	69.3	1.9	5.6	17.2	22.8	52.5
Male	17,218	71.2	3.6	9.3	20.3	22.0	44.8
Unknown	304	65.6	0.0	3.6	10.5	18.1	67.8
Grade							
7	1	58.0	0.0	0.0	0.0	0.0	100.0
8	19	66.6	0.0	5.3	5.3	31.6	57.9
9	9,262	70.8	3.6	9.0	20.1	20.6	46.7
10	22,740	71.0	2.7	7.5	20.0	24.2	45.6
11	3,659	65.8	1.0	3.5	10.5	18.3	66.7
12	1,440	64.5	1.2	2.4	7.5	16.0	72.9
Other	35	63.4	0.0	2.9	11.4	14.3	71.4
Ethnicity							
White	21,835	73.6	4.1	10.7	25.0	25.4	34.9
African American	13,561	64.8	0.4	1.9	8.7	17.8	71.3
Hispanic	712	67.6	2.2	4.4	13.2	21.5	58.7
Asian/Hawaiian-Pac. Islander	461	74.9	9.1	14.8	20.0	21.5	34.7
American Indian	59	69.3	0.0	6.8	22.0	16.9	54.2
Other	225	69.7	1.3	8.4	18.2	20.0	52.0
Unknown	303	65.4	0.0	3.0	10.2	19.1	67.7
Language							
English speaker	35,541	70.3	2.8	7.4	19.0	22.5	48.4
Exited	201	69.3	2.0	9.5	11.9	23.9	52.7
LEP mainstream	40	68.5	0.0	5.0	7.5	27.5	60.0
Full LEP	197	62.1	0.5	1.5	5.1	11.7	81.2
Waiver	18	71.7	0.0	11.1	22.2	16.7	50.0
Other	1,159	66.3	0.5	4.3	11.0	20.3	63.9
Lunch							
Free meals	10,311	65.2	0.4	2.6	9.7	17.6	69.7
Reduced-price meals	2,151	68.1	1.2	4.0	14.0	24.5	56.3
No free/reduced-price meals	23,693	72.7	3.9	9.8	23.2	24.4	38.7
Unknown	1,001	66.0	0.3	4.1	11.4	18.6	65.6
IEP							
Yes	1,843	63.0	0.7	2.0	7.7	13.1	76.5
No	1,488	69.3	2.8	7.1	16.1	19.9	54.1
Unknown	33,825	70.6	2.8	7.6	19.3	23.0	47.3
Migrant							
Yes	31	65.5	0.0	0.0	16.1	12.9	71.0
No	6,307	68.7	2.0	6.0	16.2	20.6	55.3

TABLE 5.6

Biology 1/Applied Biology 2 Implementation Test, Regular Schools: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Unknown	30,818	70.5	2.8	7.6	19.1	22.8	47.8
Courses taken							
3221 (Bio 1)	31,627	71.2	3.1	8.2	20.4	23.6	44.8
3227 (App Bio 2)	4,584	63.4	0.3	1.3	7.3	15.1	75.9
Other	945	67.1	1.3	5.9	12.3	18.8	61.7
Gifted/talented							
Academic	444	77.8	10.8	17.1	30.0	18.5	23.6
Artistic	43	69.0	2.3	4.7	18.6	25.6	48.8
Both	58	77.2	8.6	17.2	36.2	10.3	27.6
No	4,364	67.8	1.3	5.1	14.4	20.3	58.8
Unknown	32,247	70.4	2.8	7.4	19.0	22.7	48.1
504 plan							
Yes	271	69.6	2.6	5.9	19.6	19.6	52.4
No	35,155	70.4	2.8	7.5	18.9	22.6	48.3
Unknown	1,730	66.3	0.5	3.7	12.4	19.4	64.0
Alternative school							
Yes	444	64.6	0.9	3.4	10.1	15.3	70.3
No	34,468	70.4	2.8	7.5	18.9	22.6	48.1
Unknown	2,244	67.7	0.8	5.2	14.7	20.2	59.1

TABLE 5.7

Biology 1/Applied Biology 2 Implementation Test, Adult Education Programs: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Overall	77	60.8	0.0	1.3	3.9	6.5	88.3
Gender							
Female	40	58.2	0.0	0.0	0.0	2.5	97.5
Male	35	63.3	0.0	2.9	8.6	11.4	77.1
Unknown	2	67.0	0.0	0.0	0.0	0.0	100.0
Grade							
9	29	59.2	0.0	0.0	0.0	10.3	89.7
10	11	65.8	0.0	0.0	9.1	18.2	72.7
11	8	65.4	0.0	12.5	12.5	0.0	75.0
12	18	59.2	0.0	0.0	5.6	0.0	94.4
Other	11	59.0	0.0	0.0	0.0	0.0	100.0
Ethnicity							
White	35	62.5	0.0	2.9	5.7	11.4	80.0
African American	40	59.3	0.0	0.0	2.5	2.5	95.0

TABLE 5.7

Biology 1/Applied Biology 2 Implementation Test, Adult Education Programs: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Other	1	66.0	0.0	0.0	0.0	0.0	100.0
Unknown	1	53.0	0.0	0.0	0.0	0.0	100.0
Language							
English speaker	54	59.9	0.0	0.0	3.7	7.4	88.9
Other	23	62.7	0.0	4.3	4.3	4.3	87.0
Lunch							
Free meals	21	60.4	0.0	0.0	4.8	4.8	90.5
Reduced-price meals	3	61.3	0.0	0.0	0.0	33.3	66.7
No free/reduced-price meals	28	60.5	0.0	0.0	7.1	10.7	82.1
Unknown	25	61.2	0.0	4.0	0.0	0.0	96.0
IEP							
Yes	2	67.5	0.0	0.0	50.0	0.0	50.0
No	41	59.9	0.0	0.0	2.4	9.8	87.8
Unknown	34	61.4	0.0	2.9	2.9	2.9	91.2
Migrant							
No	53	60.5	0.0	0.0	5.7	9.4	84.9
Unknown	24	61.3	0.0	4.2	0.0	0.0	95.8
Courses taken							
3221 (Bio 1)	52	61.1	0.0	0.0	3.8	9.6	86.5
3227 (App Bio 2)	8	58.8	0.0	0.0	0.0	0.0	100.0
Other	17	60.8	0.0	5.9	5.9	0.0	88.2
Gifted/talented							
No	49	60.3	0.0	0.0	4.1	10.2	85.7
Unknown	28	61.6	0.0	3.6	3.6	0.0	92.9
504 plan							
Yes	1	58.0	0.0	0.0	0.0	0.0	100.0
No	41	60.6	0.0	0.0	7.3	7.3	85.4
Unknown	35	61.0	0.0	2.9	0.0	5.7	91.4
Alternative school							
Yes	40	61.2	0.0	0.0	5.0	12.5	82.5
No	10	58.7	0.0	0.0	10.0	0.0	90.0
Unknown	27	60.9	0.0	3.7	0.0	0.0	96.3

TABLE 5.8

Physical Science Implementation Test, Regular Schools:
Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Overall	45,721	65.5	2.2	4.9	10.1	16.9	65.9
Gender							
Female	22,930	65.2	1.6	4.0	9.5	17.2	67.7
Male	22,494	65.8	2.9	5.9	10.8	16.6	63.8
Unknown	297	59.8	0.3	1.3	4.4	8.4	85.5
Grade							
7	3	60.7	0.0	0.0	33.3	0.0	66.7
8	308	74.7	6.2	11.4	24.7	23.4	34.4
9	38,877	65.6	2.3	5.0	10.1	17.0	65.6
10	5,453	64.8	1.9	4.5	9.9	15.8	67.9
11	751	62.5	0.7	2.3	7.5	15.0	74.6
12	277	61.7	0.7	4.0	7.6	9.4	78.3
Other	52	59.7	1.9	0.0	5.8	13.5	78.8
Ethnicity							
White	25,365	68.8	3.6	7.4	14.1	20.6	54.3
African American	18,256	60.9	0.4	1.4	4.6	12.0	81.6
Hispanic	989	62.4	1.1	2.8	6.0	14.0	76.1
Asian/Hawaiian-Pac. Islander	438	72.2	7.5	11.9	19.6	16.2	44.7
American Indian	80	64.6	1.3	2.5	13.8	18.8	63.8
Other	285	64.3	3.5	3.5	7.7	15.4	69.8
Unknown	308	60.3	0.6	1.0	4.2	9.1	85.1
Language							
English speaker	43,769	65.6	2.3	5.0	10.3	17.1	65.4
Exited	184	64.9	2.7	4.3	7.1	19.6	66.3
LEP mainstream	71	62.5	0.0	7.0	5.6	16.9	70.4
Full LEP	305	58.0	0.3	1.3	4.3	7.5	86.6
Waiver	23	68.4	8.7	0.0	4.3	17.4	69.6
Other	1,369	62.0	1.1	2.7	6.6	12.2	77.4
Lunch							
Free meals	15,678	61.2	0.4	1.7	4.9	12.5	80.6
Reduced-price meals	3,158	64.2	1.1	3.6	8.4	17.0	69.9
No free/reduced-price meals	25,853	68.4	3.6	7.2	13.6	19.7	55.9
Unknown	1,032	60.9	0.5	2.0	5.7	11.2	80.5
IEP							
Yes	3,913	57.7	0.2	0.7	2.8	6.9	89.3
No	2,543	63.4	2.2	4.0	7.7	13.8	72.2
Unknown	39,265	66.4	2.5	5.4	11.0	18.0	63.2
Migrant							
Yes	49	58.7	0.0	2.0	6.1	6.1	85.7
Nor	8,391	64.0	2.4	4.2	8.0	14.7	70.6

TABLE 5.8

Physical Science Implementation Test, Regular Schools: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Unknown	37,281	65.8	2.2	5.1	10.6	17.4	64.8
Courses taken							
2211 (Phy Sci, grade 7 or 8)	758	71.1	7.0	8.6	17.5	18.3	48.5
3211 (Phy Sci)	43,573	65.4	2.2	4.8	10.0	16.9	66.1
Other	1,390	64.0	1.4	5.3	9.0	14.0	70.4
Gifted/talented							
Academic	527	76.1	13.3	14.2	20.3	20.1	32.1
Artistic	48	67.3	2.1	8.3	12.5	16.7	60.4
Both	33	75.1	9.1	27.3	15.2	9.1	39.4
No	6,216	62.9	1.7	3.2	6.9	14.2	74.0
Unknown	38,897	65.7	2.2	5.0	10.5	17.2	65.1
504 plan							
Yes	317	63.4	1.9	3.5	6.6	17.0	71.0
No	43,464	65.6	2.3	5.0	10.3	17.0	65.4
Unknown	1,940	62.4	0.6	2.1	6.8	13.9	76.7
Alternative school							
Yes	675	59.2	0.1	2.1	2.5	8.7	86.5
No	42,563	65.7	2.3	5.1	10.4	17.0	65.2
Unknown	2,483	63.3	1.2	2.4	7.3	16.4	72.8

TABLE 5.9

Physical Science Implementation Test, Adult Education Programs: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
Overall	96	59	0.0	1.0	3.1	5.2	90.6
Gender							
Female	37	58	0.0	0.0	2.7	0.0	97.3
Male	58	59	0.0	1.7	3.4	8.6	86.2
Unknown	1	68	0.0	0.0	0.0	0.0	100.0
Grade							
9	53	59	0.0	0.0	3.8	9.4	86.8
10	18	57	0.0	0.0	5.6	0.0	94.4
11	5	63	0.0	20.0	0.0	0.0	80.0
12	11	57	0.0	0.0	0.0	0.0	100.0
Other	9	57	0.0	0.0	0.0	0.0	100.0
Ethnicity							

TABLE 5.9

Physical Science Implementation Test, Adult Education Programs: Percentages of Student Scores in Letter-Grade Equivalents, Overall and by Demographics

Demographics	N	Mean Scale Score	A	B	C	D	F
White	46	61	0.0	2.2	6.5	4.3	87.0
African American	46	56	0.0	0.0	0.0	2.2	97.8
Hispanic	1	72	0.0	0.0	0.0	100.0	0.0
Other	3	63	0.0	0.0	0.0	33.3	66.7
Language							
English speaker	86	59	0.0	1.2	3.5	4.7	90.7
Other	10	59	0.0	0.0	0.0	10.0	90.0
Lunch							
Free meals	45	57	0.0	0.0	2.2	6.7	91.1
Reduced-price meals	1	64	0.0	0.0	0.0	0.0	100.0
No free/reduced-price meals	35	61	0.0	2.9	5.7	5.7	85.7
Unknown	15	57	0.0	0.0	0.0	0.0	100.0
IEP							
Yes	8	55	0.0	0.0	0.0	12.5	87.5
No	76	59	0.0	1.3	3.9	5.3	89.5
Unknown	12	56	0.0	0.0	0.0	0.0	100.0
Migrant							
No	84	59	0.0	1.2	3.6	6.0	89.3
Unknown	12	56	0.0	0.0	0.0	0.0	100.0
Courses taken							
3211 (Phy Sci)	93	59	0.0	1.1	3.2	5.4	90.3
Other	3	57	0.0	0.0	0.0	0.0	100.0
Gifted/talented							
Academic	2	55	0.0	0.0	0.0	0.0	100.0
No	79	59	0.0	1.3	3.8	6.3	88.6
Unknown	15	57	0.0	0.0	0.0	0.0	100.0
504 plan							
No	80	59	0.0	1.3	3.8	6.3	88.8
Unknown	16	57	0.0	0.0	0.0	0.0	100.0
Alternative school							
Yes	73	59	0.0	1.4	4.1	6.8	87.7
No	9	57	0.0	0.0	0.0	0.0	100.0
Unknown	14	57	0.0	0.0	0.0	0.0	100.0

Chapter 6

DESCRIPTIVE STATISTICS

Descriptive statistics of scale score distributions for the three 2003–04 test administrations combined are presented in table 6.1, for students overall and by gender and race. In figures 6.1 through 6.9, the scale score distributions are shown by form and are reported for students overall and by gender and race.

TABLE 6.1

2003–04 EOCEP Test Administration Summary Statistics: Regular Schools and Adult Education Programs, Overall and by Gender and Race

Algebra 1/Math Tech 2							
Regular Schools				**Adult Education Programs**			
		Scale Score				**Scale Score**	
	N	**Mean**	**SD**		**N**	**Mean**	**SD**
Overall	57,861	79.4	11.8	**Overall**	156	69.1	10.2
Gender				**Gender**			
Females	29,040	79.7	11.5	Females	71	69.8	10.1
Males	28,522	79.3	12.1	Males	80	68.9	10.3
Race				**Race**			
Whites	32,115	82.7	11.4	Whites	53	73.2	12.0
African Americans	22,932	74.9	10.8	African Americans	94	66.4	7.9
English 1							
Regular Schools				**Adult Education Programs**			
		Scale Score				**Scale Score**	
	N	**Mean**	**SD**		**N**	**Mean**	**SD**
Overall	55,902	72.3	13.0	**Overall**	108	62.5	12.0
Gender				**Gender**			
Females	27,536	73.7	12.5	Females	46	64.6	9.2
Males	27,869	71.1	13.5	Males	62	60.9	13.6
Race				**Race**			
Whites	30,830	76.2	12.7	Whites	55	63.4	12.7
African Americans	21,985	67.2	11.4	African Americans	46	61.1	10.7
Biology 1/Applied Biology 2							
Regular Schools				**Adult Education Programs**			
		Scale Score				**Scale Score**	
	N	**Mean**	**SD**		**N**	**Mean**	**SD**
Overall	37,156	70.2	11.0	**Overall**	77	60.8	8.3
Gender				**Gender**			
Females	19,634	69.3	10.3	Females	40	58.2	6.8
Males	17,218	71.2	11.6	Males	35	63.3	9.2
Race				**Race**			
Whites	21,835	73.6	10.6	Whites	35	62.5	9.5
African Americans	13,561	64.8	9.1	African Americans	40	59.3	7.0

TABLE 6.1

2003–04 EOCEP Test Administration Summary Statistics: Regular Schools and Adult Education Programs, Overall and by Gender and Race

Physical Science							
Regular Schools				**Adult Education Programs**			
		Scale Score				**Scale Score**	
	N	**Mean**	**SD**		**N**	**Mean**	**SD**
Overall	45,721	65.5	12.1	**Overall**	96	58.7	8.8
Gender				**Gender**			
Females	22,930	65.2	11.3	Females	37	57.8	7.4
Males	22,494	65.8	12.9	Males	58	59.1	9.6
Race				**Race**			
Whites	25,365	68.8	12.3	Whites	46	60.6	10.0
African Americans	18,256	60.9	10.2	African Americans	46	56.2	6.7

FIGURE 6.1

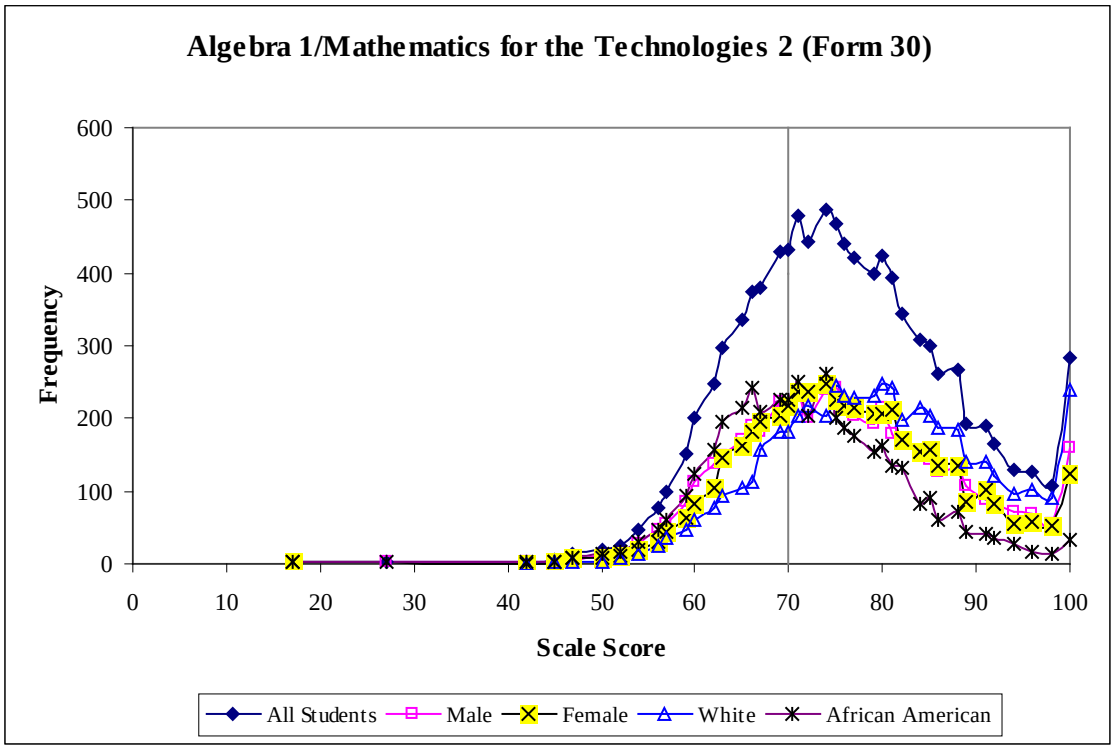


FIGURE 6.2

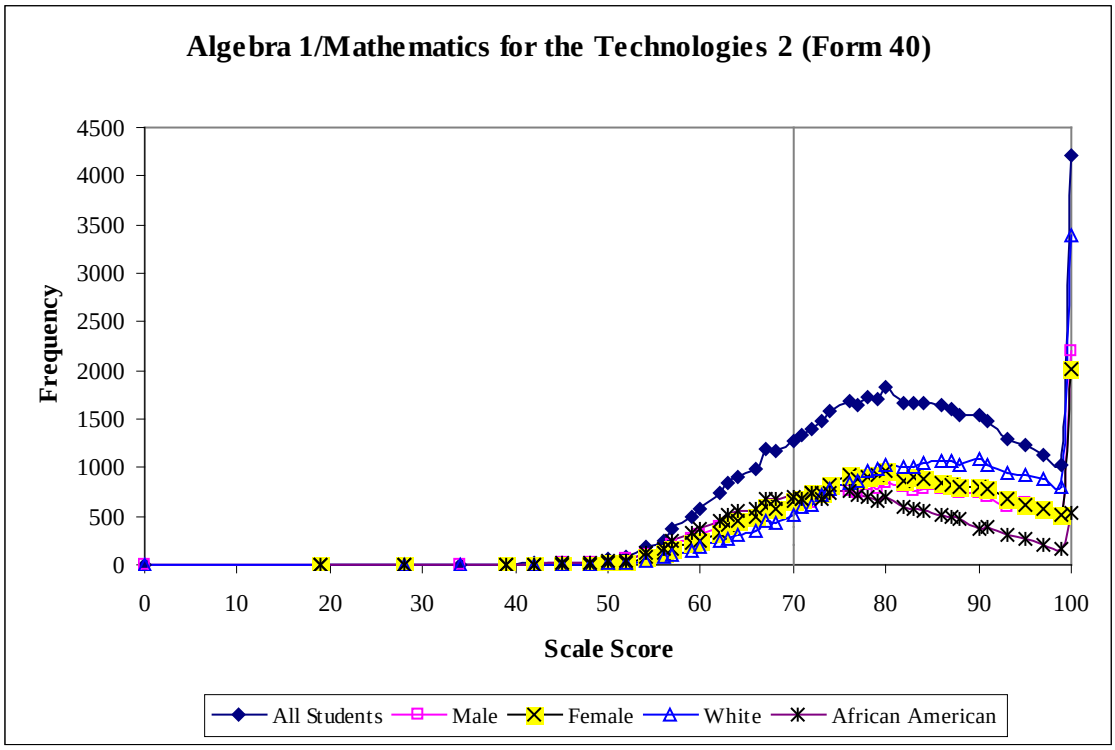


FIGURE 6.3

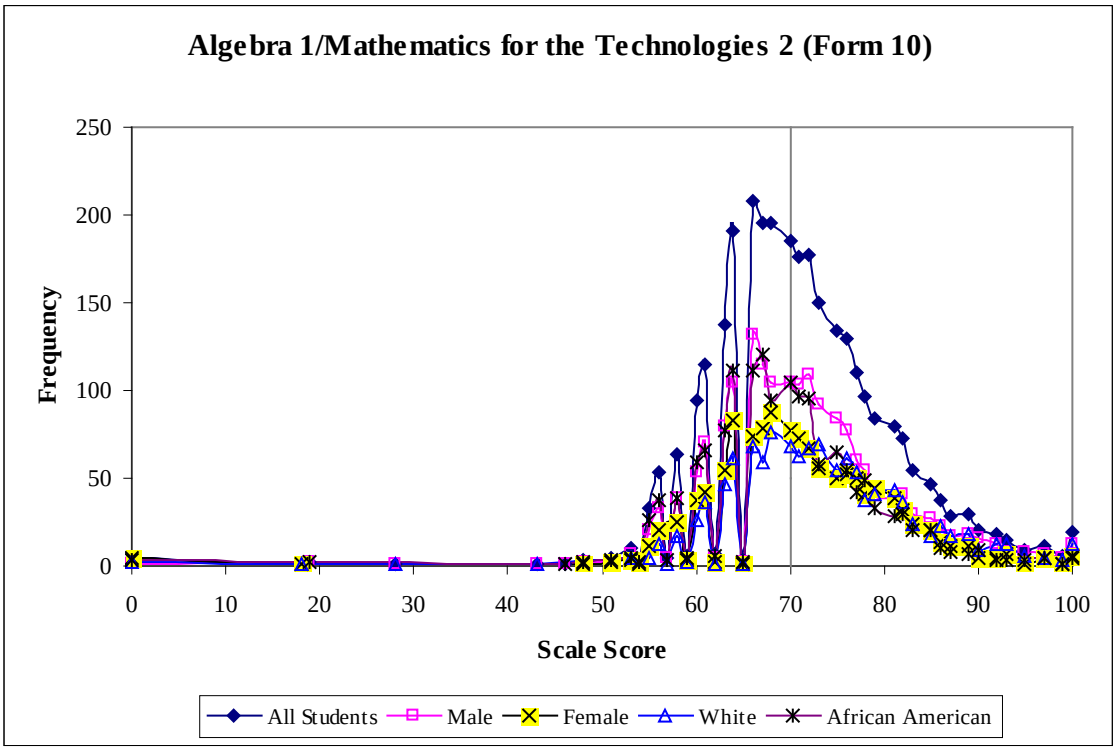


FIGURE 6.4

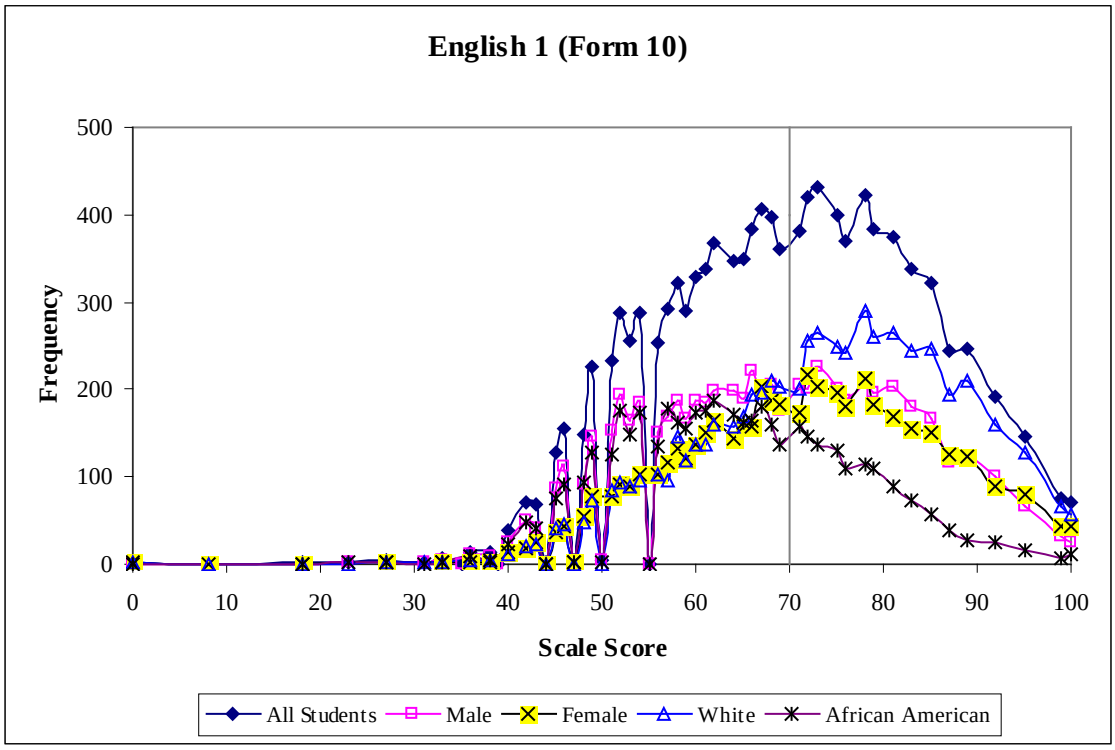


FIGURE 6.5

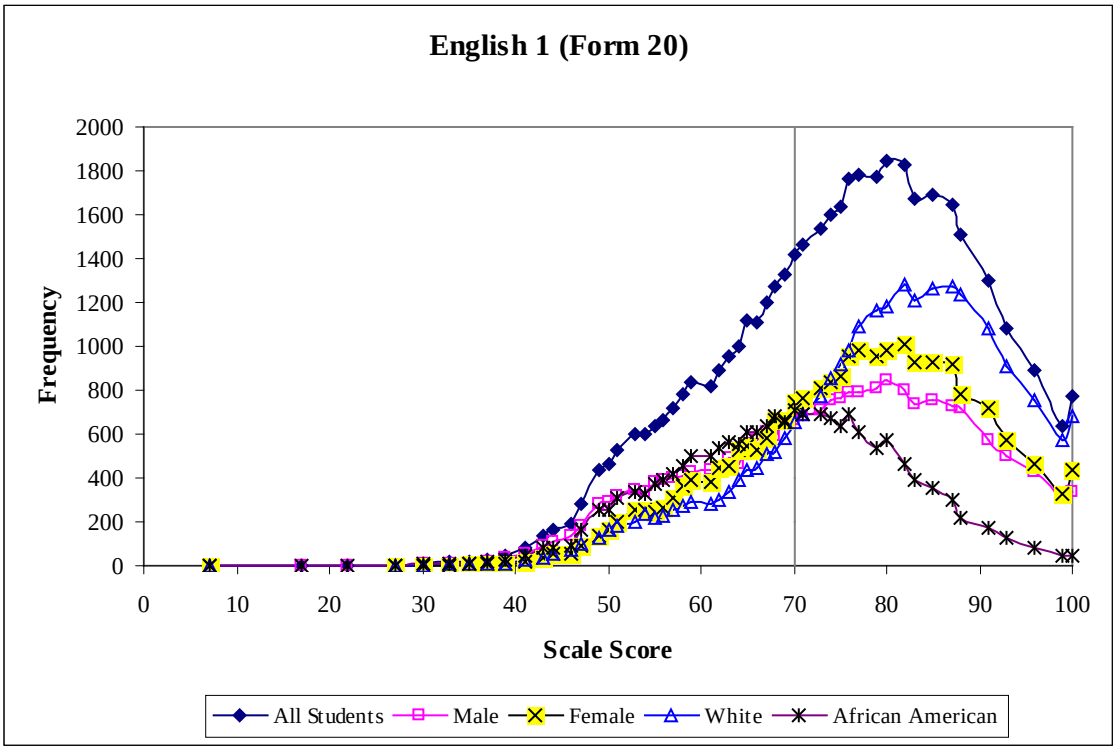


FIGURE 6.6

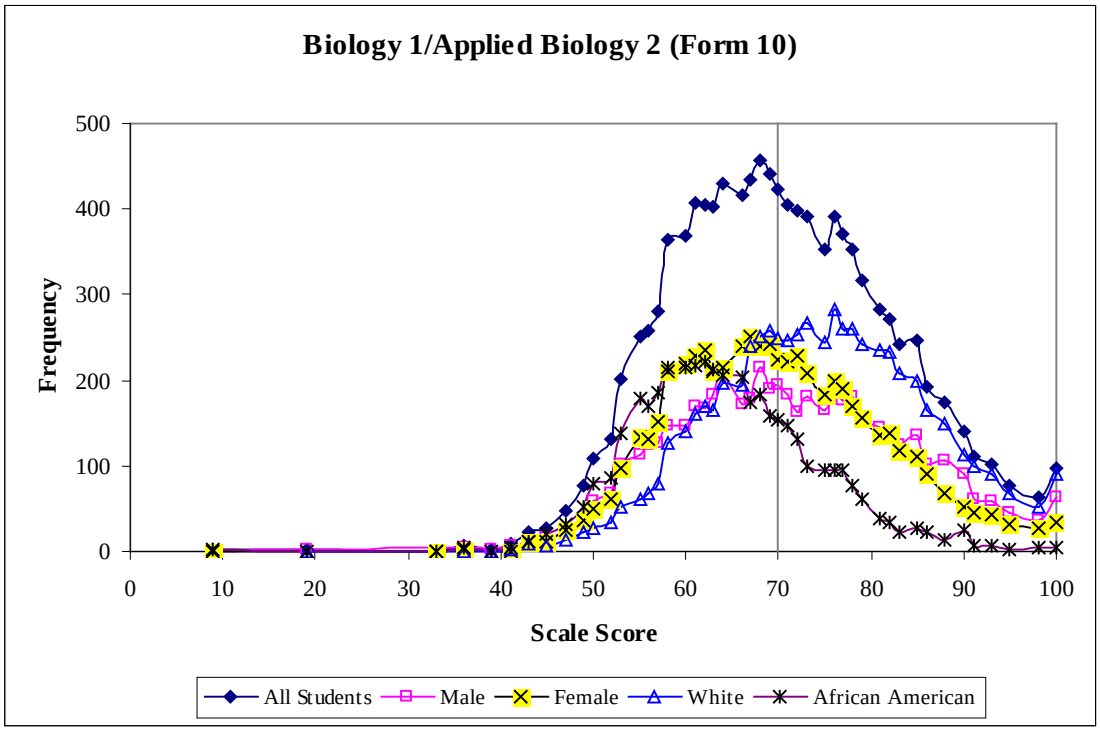


FIGURE 6.7

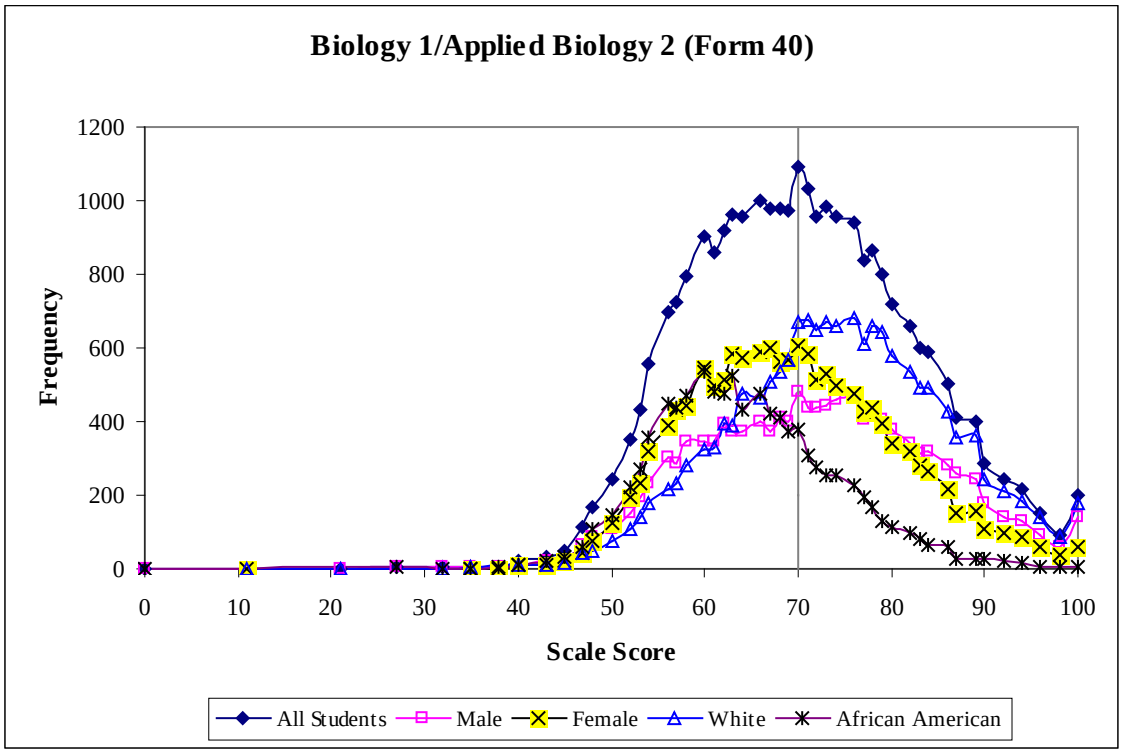


FIGURE 6.8

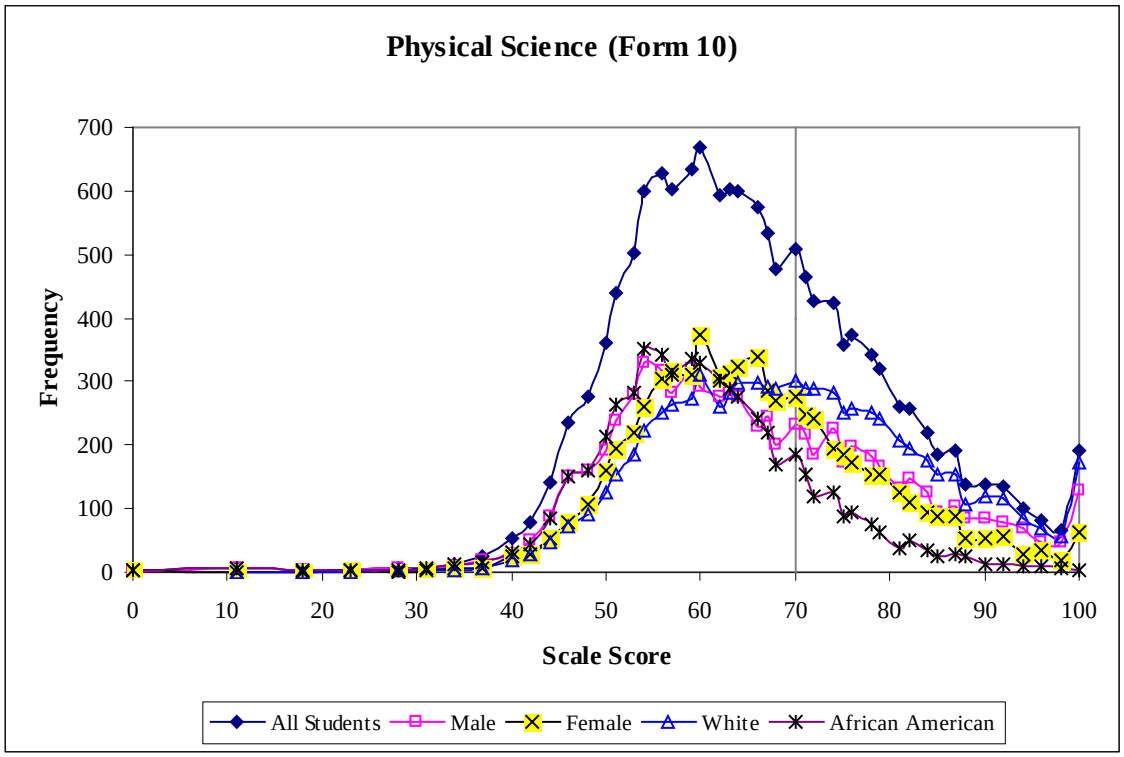
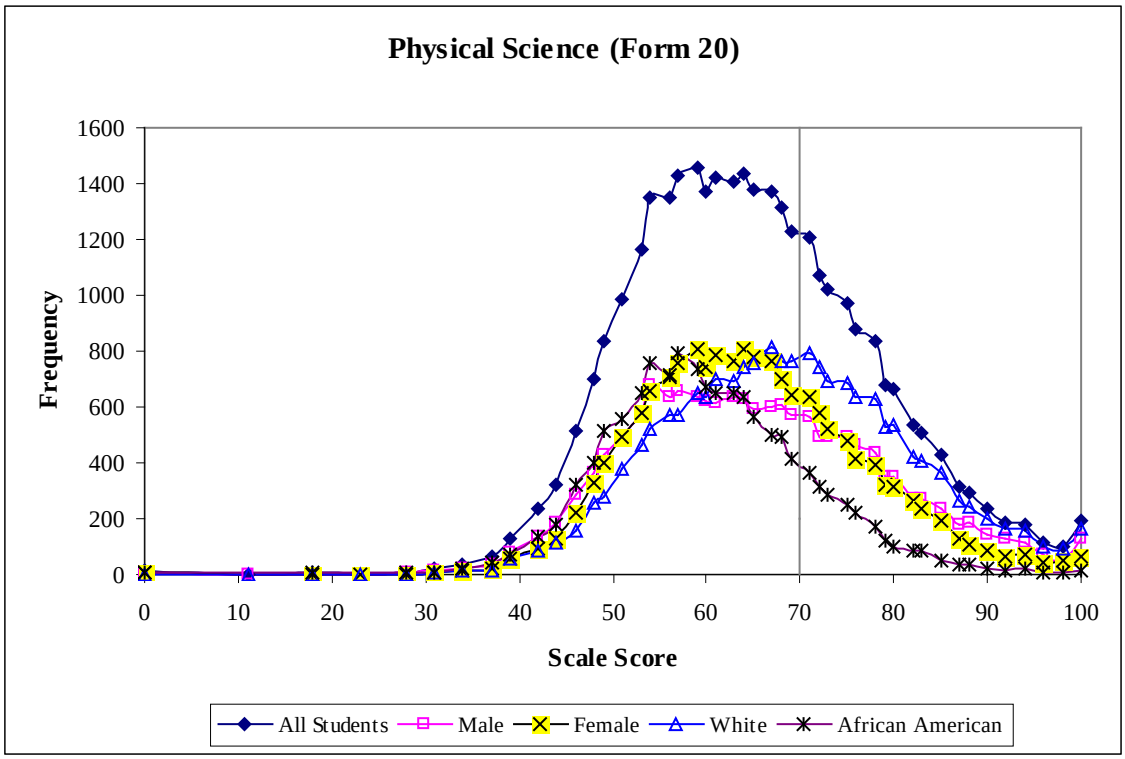


FIGURE 6.9



Chapter 7

RELIABILITY

In this chapter, multiple types of reliability indexes are presented: reliability of raw scores, overall standard error of measurement (SEM), conditional standard error of measurements for raw scores (CSEM) and scale scores (IRT SEM), and decision consistency of the passing cut score.

7.1 RELIABILITY OF RAW SCORES

Table 7.1 reports the reliability coefficients and standard errors of measurement. The reliabilities of the total raw scores were computed using the Kuder-Richardson coefficient (KR-20). The KR-21 reliability coefficients were used in computing the conditional standard error of measurement for raw scores in section 7.2.

TABLE 7.1

Reliability Coefficients of Raw Scores

Administration	Number of Items	Number of Test-Takers	KR-20	KR-21	Classical SEM
Algebra 1/Mathematics for the Technologies 2					
Fall 2003 (Form 30)	50	9,774	0.84	0.82	3.16
Spring 2004 (Form 40)	50	45,244	0.89	0.87	3.03
English 1					
Fall 2003 (Form 10)	50	8,823	0.90	0.89	3.02
Spring 2004 (Form 20)	55	44,793	0.90	0.89	3.15
Biology 1/Applied Biology 2					
Fall 2003 (Form 10)	55	10,461	0.86	0.84	3.29
Spring 2004 (Form 40)	55	26,282	0.86	0.84	3.32
Physical Science					
Fall 2003 (Form 10)	55	12,852	0.85	0.84	3.41
Spring 2004 (Form 20)	55	31,967	0.83	0.81	3.39

7.2 OVERALL AND CONDITIONAL SEM

The overall classical SEM is defined as $s_x\sqrt{1-r_{xx}}$, where s_x is the standard deviation of the scale score and r_{xx} is the reliability coefficient. The CSEM for raw scores at the cut score was computed using the following formula (Feldt and Qualls 1998; Huynh, Meyer, and Barton 2000):

$$CSEM = \sqrt{\left(\frac{1-KR20}{1-KR21}\right)\left(\frac{c(k-c)}{k-1}\right)}$$, where c = cut score and k = number of items.

The conditional standard error of the scale score (IRT CSEM) at the cut score was computed on the basis of the SEM of the Rasch ability cut score. At each ability score on the Rasch *theta* scale, the SEM is computed as the square root of the inverse of the test information function.

Table 7.2 presents the CSEM and the IRT CSEM. The CSEM at the scale score cut is defined as the reciprocal of the square root of the test information function at the point on the ability continuum that corresponds to the final scale score cut (Hambleton, Swaminathan, and Rogers 1991). Although both classical SEM and IRT CSEM serve the same role, the values of IRT-based conditional standard errors vary with ability level, whereas the classical SEM does not.

TABLE 7.2

Conditional Standard Errors of Measurement

Administration	CSEM	IRT-Based Conditional SEM
Algebra/Math Tech 2		
Fall 2003 (Form 30)	3.28	4.15
Spring 2004 (Form 40)	3.34	4.10
English 1		
Fall 2003 (Form 10)	3.22	4.05
Spring 2004 (Form 20)	3.40	4.06
Biology 1/Applied Biology 2		
Fall 2003 (Form 10)	3.46	3.95
Spring 2004 (Form 40)	3.52	3.95
Physical Science		
Fall 2003 (Form 10)	3.59	4.83
Spring 2004 (Form 20)	3.58	4.84

Note: The CSEM metric is in raw score points, and the IRT SEM metric is in scale score points.

7.3 CONSISTENCY OF PASSING CUT SCORES

When student performance is reported in a pass or fail category, a reliability index is computed in terms of the probabilities of consistent classification of students, as specified in standard 2.15 in *Standards for Educational and Psychological Testing* (AERA, NCME, and APA 1999). This index takes into consideration the consistency of classifications for the percentage of examinees who would be classified in the same way on a second (hypothetical) EOCEP administration using either the same form or an alternate equivalent form.

Although a number of procedures are available for estimating misclassification errors (Livingston and Lewis 1995; Hanson and Brennan 1990; Huynh 1976; Subkoviak 1976), the AIR used the *beta* binomial distribution method (Huynh 1979; Huynh, Meyer, and Barton 2000). Table 7.3 presents a summary of agreements between the operational test classifications; that is, the percentages of students who are consistently classified in the same category (pass or fail) on two equivalent administrations of the test. The consistency index for the passing score is computed for each administration.

TABLE 7.3

Consistency Index for Passing Scores

Administration	Consistency Index
Algebra/Math Tech 2	
Fall 2003 (Form 30)	88.7%
Spring 2004 (Form 40)	93.0%
English 1	
Fall 2003 (Form 10)	89.2%
Spring 2004 (Form 20)	90.0%
Biology 1/Applied Biology 2	
Fall 2003 (Form 10)	87.1%
Spring 2004 (Form 40)	87.3%
Physical Science	
Fall 2003 (Form 10)	87.3%
Spring 2004 (Form 20)	87.2%

Chapter 8

VALIDITY

Three types of validity evidence are reported for the algebra test forms: test content, item fairness, and internal structure. Evidence of content validity is presented in the item content distribution across domains and the alignment of the 2003–04 EOCEP test items with the state content standards. Evidence of item fairness is examined with the information on differential item functioning (DIF). Evidence of internal structure is provided in correlations among content domains.

8.1 ITEM DISTRIBUTION ACROSS CONTENT DOMAINS

The EOCEP operational and implementation test forms were constructed according to the test specifications and the test blueprints. These items measured the specific assessment standards that were approved by the SDE. All items in the test forms were reviewed by the content review committee and the sensitivity review committee and were approved by the SDE. The 2003–04 EOCEP test form specifications are presented in tables 8.1 through 8.4 by subject.

TABLE 8.1

Item Distribution by Content Domain for Algebra 1/Math Tech 2

Content Domain	Form 10	Form 30	Form 40
I. Understanding Functions			
A. Relationships	5	5	5
B. Linear and Quadratic Functions and Data Representations	5	5	5
C. Generalizations, Algebraic Symbols, and Matrices	4	4	4
D. Algebraic Expressions in Problem Solving Situations	6	6	6
II. Linear Functions			
A. Representations	4	4	4
B. Interpretations	8	8	8
C. Equations and Inequalities	7	8	7
D. Systems of Linear Equations	3	2	3
III. Quadratic and Other Functions			
A. Quadratic Functions	5	5	5
B. Other Functions	3	3	3
Totals	**50**	**50**	**50**

TABLE 8.2

Item Distribution by Content Domain for English 1

Strand/Topic	Form 10	Form 20
Reading Process and Comprehension	12	13
Analysis of Texts	13	14
Word Study and Analysis	8	9
Research	4	3
Writing	8	11
Communication	5	5
Totals	**50**	**55**

TABLE 8.3

Item Distribution by Content Domain for Biology 1/Applied Biology 2

Content Domain	Form 10	Form 40
I. Inquiry		
A. Identify Questions	2	2
B. Design and Conduct Investigations	10	11
C. Math in Science	1	1
D. Scientific Explanations	1	0
E. Alternative Explanations	0	0
F. Scientific Inquiry	0	0
II. Biology		
A. The Cell	9	9
B. Heredity	7	7
C. Biological Evolution	5	5
D. Interdependence of Organisms	9	9
E. Matter, Energy and Organization	7	7
F. Behavior and Regulation	4	4
Totals	**55**	**55**

TABLE 8.4

Item Distribution by Content Domain for Physical Science

Content Domain	Form 10	Form 20
I. Inquiry		
A. Identify Questions	2	2
B. Design and Conduct Investigations	6	6
C. Math in Science	4	4
D. Scientific Explanations	1	1
E. Alternative Explanations	0	0
G. Scientific Inquiry	0	1
II. Chemistry		
A. Structure of Atoms	3	3
B. Structure and Properties of Matter	11	11
C. Chemical Reactions	6	6
III. Physics		
A. Motions and Forces	13	12
B. Conservation of Energy and the Increase in Disorder	5	5
C. Interactions of Energy and Matter	4	4
Totals	**55**	**55**

8.2 ITEM DEVELOPMENT

All EOCEP items were developed with reference to the South Carolina curriculum standards and measurement guidelines. Various committees reviewed all items; only items approved by these committees and the SDE were included in the operational forms.

The field-test items in the fall 2003 and spring 2004 administrations were also thoroughly reviewed before being included in the operational forms. The AIR reviewed the field-test items internally before sending them to the SDE for review.

8.3 DIFFERENTIAL ITEM FUNCTIONING

A critical issue in statewide high-stakes testing is whether the test is “fair” to all test-takers; therefore, an important goal of item and test development is to produce a pool of items that are judged to be free of bias, either toward or against, any group of students. All EOCEP items were reviewed both for bias and for DIF. The sensitivity review committee examined the EOCEP items for potential bias, including language that might disadvantage a particular group, might be considered offensive to members of a particular group, or might present obstacles to a particular group due to factors unrelated to content and processes specified in the standards. After data were collected, the DIF statistics were produced to examine the items statistically.

A psychometric definition of the term "test fairness" is the degree to which an item performs differently for one group of examinees than it performs for another group of equally able examinees. The term "differential item functioning" (DIF) refers to the statistical properties of an item in two equally able groups, which are subject to later interpretation and judgment by measurement and content experts. Once an item is flagged for a significant DIF, judgment is used to determine whether the difference in difficulty shown by the DIF index is unfairly related to group membership. The DIF statistics should be seen not as indicators of bias, or unfairness, but as indicators of relative strengths and weaknesses of the two groups being compared when the overall ability that the test is intended to measure has been controlled.

As with other statistical methodologies, there are numerous widely accepted approaches to detecting potential unfairness in test items. Many of these methods fall under the general category of DIF analyses.

Procedure:

The procedure that the AIR selected for detecting DIF was the Mantel-Haenszel (MH) chi-square for dichotomous items. The AIR calculated the Mantel-Haenszel statistic (MH D-DIF) for MC items (Holland and Thayer 1988) to measure the degree and magnitude of DIF. The examinee group of interest is the *focal* group, and the group to which performance on the item is being compared is the *reference* group. In this report, the focal groups for DIF were females and African Americans.

Items were separated into one of three categories on the basis of DIF statistics (Holland and Thayer 1988; Dorans and Holland 1993): negligible DIF (category A), intermediate DIF (category B), and large DIF (category C). The items in category C, which exhibit significant DIF, are of primary concern.

Positive values of *delta* indicate that the item is easier for the *focal* group, suggesting that the item favors the *focal* group. A negative value of *delta* indicates that the item is more difficult for the *focal* group. The item classifications are based on the Mantel-Haenszel chi-square and the MH delta (Δ) value as follows:

- The item is classified as C category if the absolute value of the MH delta value (i.e., $|\Delta|$) is significantly greater than 1 and also greater than or equal to 1.5.
- The item is classified as B category if the MH delta value (Δ) is significantly different from 0 and either the absolute value of the MH delta ($|\Delta|$) is less than 1.5 or the absolute value of the MH delta ($|\Delta)|$ is not significantly different from 1.
- The item is classified as A category if delta value (Δ) is not significantly different from 0 or the absolute value of delta ($|\Delta|$) is less than or equal to 1.

The data in tables 8.5 and 8.6 summarize the number of items in DIF categories for the 2003–04 test operational and field-test items.

When the operational forms were constructed, all item statistics from the initial field test were reviewed and approved by the SDE. Inclusion of any "flagged" items on an operational form (i.e., items classified as C category) was possible only when the SDE approved the inclusion of a particular item.

In the fall 2003 test forms, no significant DIF items were included in the operational forms. When the operational test data were analyzed, two items showed a significant DIF in the Algebra 1/Mathematics for the Technologies 2 test (C– in ethnicity) while one item in English 1 exhibited C– in gender.

In the spring 2004 test forms, a few items with a significant DIF (two items in the Algebra 1/Mathematics for the Technologies 2 form and two in the English 1 form; three items in the Biology 1/Applied Biology 2 form; and no significant DIF item in the Physical Science form) were included in the test forms based on the field-test data. When the operational test data were analyzed, these items did not exhibit a significant DIF in the Biology 1/Applied Biology 2 test. In Algebra 1/Mathematics for the Technologies 2, two items—one identified through field testing as C– and one new item with C– in ethnicity—showed a significant DIF. In English 1, one item with C– in gender disappeared, but one new item showed C– in ethnicity. In Physical Science, no item showed a significant DIF.

TABLE 8.5

Summary of Differential Item Functioning for Operational Items

Administration	Cat	Whites/African Americans				Males/Females			
		Alg	Eng	Bio	PS	Alg	Eng	Bio	PS
Fall 2003	A+	22	21	26	26	27	25	24	27
	A–	26	29	27	28	21	23	27	27
	B+	0	0	0	0	1	0	3	0
	B–	0	0	2	1	1	1	1	1
	C+	0	0	0	0	0	0	0	0
	C–	2	0	0	0	0	1	0	0
Spring 2004	A+	25	29	27	25	27	31	30	27
	A–	19	21	27	30	23	20	24	28
	B+	0	1	0	0	0	2	0	0
	B–	4	3	1	0	0	2	1	0
	C+	0	0	0	0	0	0	0	0
	C–	2	1	0	0	0	0	0	0

TABLE 8.6

Summary of Differential Item Functioning for Field-Test Items

Administration	Cat	Whites/African Americans			Males/Females		
		Eng	Bio	PS	Eng	Bio	PS
Fall 2003	A+	27	4	3	20	4	3
	A–	14	1	2	19	1	2
	B+	1	0	0	2	0	0
	B–	0	0	0	1	0	0
	C+	0	0	0	0	0	0
	C–	0	0	0	0	0	0
Spring 2004	A+	53	NA	4	75	NA	3
	A–	65	NA	1	33	NA	2
	B+	1	NA	0	11	NA	0
	B–	1	NA	0	0	NA	0
	C+	0	NA	0	1	NA	0
	C–	0	NA	0	0	NA	0

8.4 CORRELATIONS AMONG CONTENT DOMAINS

Evidence of internal structure was examined in correlations among content domains. Tables 8.7 through 8.9 report mean, standard deviation, and the correlation matrices for the raw scores among content domains for Algebra 1/Mathematics for the Technologies 2, English 1, Biology 1/Applied Biology 2, and Physical Science.

TABLE 8.7

Correlations among Domain Scores for Algebra 1/Math Tech 2

Domain*	UF	LF	QOF	Number of Items
	Fall 2003 (N = 9,770)			
UF	1.00	0.67	0.49	20
LF	—	1.00	0.52	22
QOF	—	—	1.00	8
	Spring 2004 (N = 45,222)			
UF	1.00	0.75	0.62	20
LF	—	1.00	0.65	22
QOF	—	—	1.00	8

UF = Understanding Functions; LF = Linear Functions;
QOF = Quadratic and Other Functions

TABLE 8.8

Correlations among Domain Scores for English 1

Domain*	R1	R2	R3	RS	C1	W1	Number of Items
				Fall 2003 (N = 8,823)			
R1	1.00	0.68	0.68	0.49	0.46	0.50	12
R2	—	1.00	0.72	0.51	0.48	0.54	13
R3	—	—	1.00	0.49	0.47	0.53	8
RS	—	—	—	1.00	0.36	0.38	4
C1	—	—	—	—	1.00	0.43	5
W1	—	—	—	—	—	1.00	8
				Spring 2004 (N = 44,788)			
R1	1.00	0.67	0.64	0.48	0.53	0.60	13
R2	—	1.00	0.65	0.52	0.56	0.64	14
R3	—	—	1.00	0.45	0.51	0.57	9
RS	—	—	—	1.00	0.43	0.50	3
C1	—	—	—	—	1.00	0.55	5
W1	—	—	—	—	—	1.00	11

*R1, Reading Comprehension; R2, Analysis of Texts; R3, Word Analysis; RS, Research; W1, Writing; C1, Communication

TABLE 8.9

Correlations among Domain Scores for Biology 1/Applied Biology 2 and Physical Science

Biology 1/Applied Biology 2			
Domain	**Inquiry**	**Biology**	**Number of Items**
		Fall 2003 (N = 10,461)	
Inquiry	1.00	0.72	14
Biology	—	1.00	41
		Spring 2004 (N = 26,278)	
Inquiry	1.00	0.64	14
Biology	—	1.00	41

Physical Science				
Domain	**Inquiry**	**Chemistry**	**Physical Science**	**Number of Items**
		Fall 2003 (N = 12,852)		
Inquiry	1.00	0.57	0.60	13
Chemistry	—	1.00	0.63	20
Physics	—	—	1.00	22
		Spring 2004 (N = 31,965)		
Inquiry	1.00	0.55	0.54	14
Chemistry	—	1.00	0.64	20
Physics	—	—	1.00	21

WORKS CITED

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

Dorans, Neil J., and Paul W. Holland. 1993. "DIF Detection and Description: Mantel-Haenszel and Standardization." In *Differential Item Functioning*, edited by Paul W. Holland and Howard Wainer. Hillsdale, NJ: Lawrence Erlbaum Associates.

Feldt, Leonard S., and Audrey L. Qualls. 1998. "Approximating Scale Score Standard Error of Measurement from the Raw Score Standard Error." *Applied Measurement in Education* 11, no. 2:159–77.

Hambleton, Ronald K., Hariharan Swaminathan, and H. Jane Rogers. 1991. *Fundamentals of Item Response Theory.* Newbury Park, CA: Sage Publications.

Hanson, Bradley A., and Robert L. Brennan. 1990. "An Investigation of Classification Consistency Indexes Estimated under Alternative Strong True Score Models." *Journal of Educational Measurement* 27:345–59.

Holland, P. W., and Dorothy T. Thayer. 1988. "Differential Item Performance and the Mantel-Haenszel Procedure." In *Test Validity*, edited by Howard Wainer and Henry I. Braun. Hillsdale, NJ: Lawrence Erlbaum Associates.

Huynh, Huynh. 1976. "On the Reliability of Decisions in Domain-Referenced Testing." *Journal of Educational Measurement* 13:253–64.

———. 1979. "Computational and Statistical Inference for Two Reliability Indices Based on the Beta-Binomial Model." *Journal of Educational Statistics* 4:231–46.

Huynh, H., J. Patrick Meyer III, and Karen Barton. 2000. *Technical Documentation for the 1999 Palmetto Achievement Challenge Tests of English Language Arts and Mathematics, Grades Three through Eight.* Columbia: South Carolina Department of Education.

Linacre, John M., and Benjamin D. Wright. 2003. *WINSTEPS Rasch-Model Computer Program*. Chicago: Mesa Press.

Livingston, Samuel A., and Charles Lewis. 1995. "Estimating the Consistency and Accuracy of Classifications Based on Test Scores." *Journal of Educational Measurement* 32:179–97.

Rasch, Georg. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests.* Copenhagen: Danish Institute for Educational Research.

SDE. 2003a. *Fall 2003 Test Administration Manual for School Test Coordinators and Test Administrators*. Columbia: South Carolina Department of Education.

———. 2003b. *Fall 2003 District Test Coordinator's Supplement.* Columbia: South Carolina Department of Education.

———. 2004a. *Spring 2004 Test Administration Manual for School Test Coordinators and Test Administrators.* Columbia: South Carolina Department of Education.

———. 2004b. *Spring 2004 District Test Coordinator's Supplement.* Columbia: South Carolina Department of Education.

———. 2004c. *Summer 2004 Test Administration Manual for School Test Coordinators and Test Administrators.* Columbia: South Carolina Department of Education.

———. 2004d. *Summer 2004 District Test Coordinator's Supplement.* Columbia: South Carolina Department of Education.

Subkoviak, Michael J. 1976. "Estimating Reliability from a Single Administration of a Criterion-Referenced Test." *Journal of Educational Measurement* 13:265–76.

Wright, Benjamin D., and Mark H. Stone. 1979. *Best Test Design*. Chicago: Mesa Press.