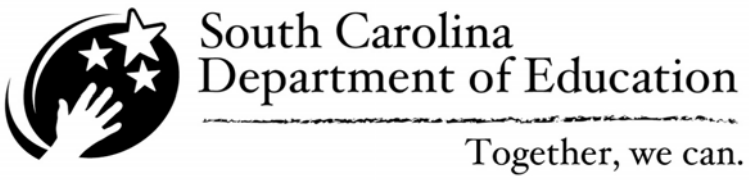
South Carolina
High School Assessment Program

English Language Arts and Mathematics
2004–05 Operational Test Technical Report



Prepared by

Bokhee Yoon, Kyunghee Suh, and Karen Thornton

American Institutes for Research

Submitted to the

South Carolina Department of Education

Inez M. Tenenbaum
State Superintendent of Education

CONTENTS

LISTS OF TABLES AND FIGURES

TABLES

FIGURES

Chapter 1

HSAP HISTORY AND OVERVIEW

The South Carolina Education Accountability Act (EAA) of 1998 mandates that all public school students pass an exit examination as one requirement for earning a high school diploma. The federal No Child Left Behind Act (NCLBA) of 2001 mandates that all states assess their public high school students' academic achievement in reading, language arts, and mathematics. The High School Assessment Program (HSAP) tests were developed to meet both of these statutory requirements by serving both as a criterion for a student's eligibility to receive a South Carolina high school diploma and as a primary source for reporting the required NCLBA data.

The HSAP tests were field-tested in spring 2003 to produce a sufficient number of items to build pre-equated operational test forms for both mathematics and English language arts (ELA). The first operational test was administered in spring 2004; the second and third operational tests were administered in fall 2004 and spring 2005, respectively.

1.1 TEST PARTICIPATION REQUIREMENTS

Student participation guidelines for HSAP testing differed for each operational administration:

- For the spring 2004 HSAP administration, all students who were enrolled in their second year of high school after their initial enrollment in the ninth grade were required to take the HSAP mathematics and ELA tests.
- For the fall 2004 administration, two groups of students were required to participate in HSAP testing: those students who had taken the tests in spring 2004 but did not pass and those who had been eligible to take the tests in spring 2004 but for a valid reason did not do so.
- For the spring 2005 administration, participation was required of all students who
  (a) were now in their second year of high school following their initial enrollment in the ninth grade,
  (b) had already taken the HSAP but who did not pass one or both of the HSAP tests,
  (c) had been eligible to take the HSAP tests in spring or fall 2004 but did not do so, or
  (d) had skipped the ninth grade and were now in their second year of high school.

  Also required to participate the spring 2005 HSAP administration were students who were beyond their second year of high school following their initial enrollment in the ninth grade but who were ineligible for Basic Skills Assessment Program (BSAP) testing—for example, a student who was now enrolled in the ninth grade for the third time or a student who had been in the ninth grade for two years and was now in the tenth grade for the first time. The exception to this requirement was any twelfth-grade student whose individualized education program (IEP) specifically stated that he or she was not to participate in the BSAP, the expectation being that such a student would complete high school prior to spring 2006, when the HSAP requirement for graduation was to go into effect.

  Students who were graduating in spring 2005 and were qualified for Graduation Express (the scoring system used for graduating seniors) reporting were also required to take the spring

2005 HSAP. Although students who qualify for Graduation Express reporting are typically in the twelfth grade, this population may include students from other grades who may have completed all Carnegie credits to qualify for early graduation. In the spring 2005 HSAP administration, test booklets for Graduation Express students were submitted for scoring. A total of 312 students who had taken the HSAP mathematics test and 164 students who taken the ELA test used the Graduation Express service, which provides HSAP results for them sooner than non-Graduation Express students are given their scores.

1.2 TEST DESIGN AND STRUCTURE

As with the spring HSAP 2004 administration, two field-test designs were combined for use in the spring 2005 ELA administration in order to increase the number of items in the item pool for some academic standards.

The first design added field-test items to the operational test form: a total of eight sets of 10 items were appended to the base form, resulting in the administration of a total of eight field test forms and 80 field-test items. These added field-test items were not used in calculating student scores.

The second design was an "embedded operational" field test, which was implemented in order for the spring 2005 ELA assessment to meet the test blueprint requirements. This approach was the administering of field-test items as part of the base (i.e., operational) test form: 7 multiple-choice items were embedded in the base form, all of which had passed qualitative and quantitative criteria and were used in calculating student scores.

Consequently, each of the eight spring 2005 ELA forms consisted of 57 operational multiple-choice items, 2 operational constructed-response items, and 1 operational extended-response item, plus the 7 embedded multiple-choice field-test items and 10 additional multiple-choice field-test items. The 60 operational and 7 embedded field-test items were common to all forms; however, the 10 additional field-test items were unique to each of the eight forms. For the score reports, 63 items were used to generate student ELA scores. This number was the result of following process: due to the fact that the 7 embedded items were determined to be qualitatively superior to 4 items that ultimately were not scored, 4 was subtracted from 60 (the initial number of operational items), and 7 was then added.

1.3 TECHNICAL REPORT CONTENT

This technical report summarizes the results of statistical and psychometric analyses performed on the fall 2004 and the spring 2005 operational data for the HSAP mathematics and ELA tests. All statistics are based on students in the regular schools only; students in adult education and district-approved homeschools are excluded. For fall 2004, the data summary in all chapters of this technical report includes students who attempted the HSAP tests for the first and second times. For spring 2005, the data in chapter 2 include students who attempted the HSAP tests for the first, second, or third time. The data in other chapters include only those students who were attempting the HSAP tests for the first time.

Chapter 2

STUDENT DEMOGRAPHICS

2.1 STUDENT PARTICIPATION

In the fall 2004 and the spring 2005 HSAP administrations, demographic data were collected for each student. These data include the categories of gender, race/ethnicity, grade, language fluency (i.e., LEP—limited English proficiency), lunch program eligibility, disability status, and migrant status.

Tables 2.1 and 2.2, respectively, report the fall 2004 and spring 2005 HSAP administrations by student demographic variables. The "Invalid" category in these tables includes blanks and multiple marks. The spring 2004 student data file, containing only students who did not pass the HSAP tests in spring 2004, was used as a fall 2004 pre-ID file. Because those students changed grade level from spring 2004 to fall 2005, the grade field was taken from the hand-gridded information. The high *invalid* rate for the "Grade" category was caused by the fact that some students and test administrators had not filled out the grade field.

TABLE 2.1

Fall 2004 Summary of Student Demographics in the Sample (All Attempts)

Demographic Category	Mathematics		ELA	
	N	%	N	%
All Students	11,373		9,112	
Gender				
Female	5,099	44.8	3,512	38.5
Male	5,977	52.6	5,335	58.6
Invalid	297	2.6	265	2.9
Ethnicity				
African American	6,969	61.3	5,430	59.6
Asian/Pacific Islander	70	0.6	91	1.0
Hispanic	334	2.9	385	4.2
American Indian	20	0.2	16	0.2
White	3,566	31.4	2,838	31.1
Other	138	1.2	113	1.2
Invalid	276	2.4	239	2.6
Grade				
09	775	6.8	675	7.4
10	1,903	16.7	1,544	16.9
11	4,029	35.4	3,471	38.1
12	251	2.2	214	2.3
Invalid	4,415	38.8	3,208	35.2
Language				
English speaker	11,188	98.4	8,871	97.4
Full LEP	119	1.1	166	1.8
LEP mainstream	35	0.3	42	0.5

TABLE 2.1

Fall 2004 Summary of Student Demographics in the Sample (All Attempts)

Demographic Category	Mathematics		ELA	
	N	%	N	%
Waiver	10	0.1	11	0.1
Exited	21	0.2	22	0.2
Lunch Program				
No free/reduced lunch	7,751	68.2	2,819	30.9
Free lunch	3,239	28.5	311	3.4
Reduced lunch	383	3.4	5,982	65.6
IEP				
No	8,852	77.8	6,773	74.3
Yes	2,512	22.1	2,334	25.6
Unknown	9	0.1	5	0.1
Migrant				
No	11,356	99.9	9,092	99.8
Yes	17	0.1	20	0.2
Unknown	—	—	—	—
Attempt				
1st	2,465	21.7	2,361	25.9
2nd	8,908	78.3	6,751	74.1

Table 2.2

Spring 2005 Summary of Student Demographics in the Sample (All Attempts)

Demographics	Mathematics		ELA	
	N	%	N	%
All Students	58,366		56,211	
Gender				
Female	28,767	49.3	27,378	48.7
Male	29,146	49.9	28,452	50.6
Invalid	453	0.8	381	0.7
Ethnicity				
African American	25,126	43.0	23,587	42.0
Asian/Pacific Islander	613	1.1	631	1.1
Hispanic	1,447	2.5	1,466	2.6
American Indian	105	0.2	107	0.2
White	30,079	51.5	29,515	52.5
Other	498	0.9	480	0.8
Invalid	498	0.9	425	0.8
Grade				
09	8,255	14.1	8,120	14.5
10	44,786	76.7	44,221	78.7
11	4,347	7.5	3,088	5.5
12	374	0.6	265	0.5

Table 2.2

Spring 2005 Summary of Student Demographics in the Sample (All Attempts)

Demographics	Mathematics		ELA	
	N	%	N	%
Invalid	604	1.0	517	0.9
Language				
English speaker	57,362	98.3	55,147	98.1
Full LEP	520	0.9	581	1.0
LEP mainstream	185	0.3	184	0.3
Waiver	55	0.1	60	0.1
Exited	243	0.4	238	0.4
Unknown	1	0.0	1	0.0
Lunch Program				
No free/reduced lunch	32,411	55.5	31,499	56.0
Free lunch	22,122	37.9	21,064	37.5
Reduced lunch	3,833	6.6	3,648	6.5
IEP				
No	49,041	84.0	4,7351	84.2
Yes	9,274	15.9	8,811	15.7
Unknown	51	0.1	49	0.1
Migrant				
No	58,315	99.9	561,58	99.9
Yes	51	0.1	53	0.1
Attempt				
1st	51,270	87.84	51,459	91.6
2nd	3,971	6.80	2,838	5.1
3rd	3,125	5.35	1,914	3.4

2.2 ACCOMMODATIONS AND MODIFICATIONS

Supplemental information regarding the administration of the HSAP to students with disabilities is provided in the *HSAP Test Administration Manual* (*TAM*) (SDE 2004a and 2005a). The *TAM* provides guidelines for IEP teams in making decisions about testing students with disabilities; it also outlines specific information regarding testing accommodations, testing modifications, test forms and materials, and administration procedures. A student with a documented disability either is one who has been evaluated and found to meet the eligibility criteria for enrollment in special education as defined by the Individuals with Disabilities Education Act of 1997 and State Board of Education Regulation 43-243.1 or is 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 HSAP assessments. Decisions about accommodations, modifications, and alternate assessment must be made on an individual student basis and not on the basis of the category of disability.

Accommodations

The term *accommodation* refers to a change in the testing environment, procedures, or presentation that does not alter what the test measures or the comparability of scores. The purpose of accommodations is to enable students to participate in an assessment in a way that allows knowledge and skills, rather than disabilities, to be assessed.

Examples of the accommodations that were allowed during the 2004–05 HSAP administrations are changes in the test setting, timing, and scheduling: students were allowed to take the test in a different setting, such as individually or in a small group, as opposed to taking it with their class; students were allowed extended amounts of time to complete the test; and students were allowed to take the test over several days or periods during the day with frequent breaks. These are all general types of accommodations, and they can vary widely from child to child, according to what is specified in the IEP. Other accommodations allowed include the use of a poor speller's dictionary (e.g., *The Misspeller's Dictionary*) for the ELA test, oral and signed administrations of the mathematics test, and the use of customized test materials (see section 3.4 below for more details) such as loose-leaf test booklets, large-print test booklets, and braille for both tests.

Tables 2.3 and 2.4 present summaries of accommodations by the percentages of those students who were administered the test with one or more accommodations. (The column percentage totals exceed 100 percent because some students received accommodations in more than one category.)

TABLE 2.3

Percentages of Students with Accommodations in Fall 2004 HSAP Administration (All Attempts)

Accommodation	Mathematics		ELA	
	Form 31A (N = 9,212)	Customized Form (N = 2,161)	Form 31A (N = 7,012)	Customized Form (N = 2,100)
Setting	7.6	65.3		66.1
Presentation	0.6	21.3	0.8	29.7
Timing	1.0	8.4	1.1	8.9
Schedule	0.3	5.4	0.3	6.0
Spelling	—	—	0.3	5.8
Response options	0.2	5.3	0.3	9.3
Oral administration	—	72.5	—	—
Signed administration	—	1.1	—	—

TABLE 2.4

Percentages of Students with Accommodations in Spring 2005 HSAP Administration (All Attempts)

Accommodation	Mathematics		ELA	
	Form 32A (N = 54,731)	Customized Form (N = 3,635)	Form 32A (N = 51,525))	Customized Form (N = 4,686)
Setting	3.4	72.1	3.3	64.8
Presentation	—	—	0.1	0.7
Timing	0.4	13.8	0.4	13.0
Schedule	0.1	7.5	0.1	6.9
Response options	0.0	5.1	0.3	8.1
Loose leaf	—	3.5	—	2.4
Large print	—	1.8	—	1.3
Spelling	—	—	0.3	8.8
Audiocassette	—	43.5	—	—
Oral script	—	57.1	—	—
Signed administration	—	1.1	—	—
Braille	—	0.2	—	0.2
Others	0.1	0.5	—	6.9

Modifications

The term *modification* refers to a change in the testing environment, procedures, or presentation that compromise the test validity and may alter the meaning and comparability of test scores. Modifications are appropriate only for those students with disabilities who, owing to the nature of their disabilities, are unable to take the HSAP tests without modifications. The testing modifications should be the same as the modifications used by the student in routine instruction and assessment.

The 2004–05 administrations of the ELA test incorporated all of the modifications that the State Department of Education (SDE) had approved—for example, oral administration, signed administration, alternative scoring for extended-response writing items, and extended-response writing options (e.g., spell checker, grammar checker). The alternative scoring rubric was slightly different from the regular scoring rubric. If an alternative scoring accommodation was marked on a student's answer document, the extended-response writing was to be scored using the alternative scoring rubric. If a student was allowed a test modification, the modification was noted on the roster reports provided to the schools and districts and on the individual score reports. The summary results include scores for students who used modifications. Table 2.5 presents summaries of modifications by percentages. (Again, the column percentage totals exceed 100 percent because some students received modifications in more than one category.)

TABLE 2.5

Percentages of Students with Modifications in 2004–05 HSAP Administrations (All Attempts)

Fall 2004			Spring 2005		
	Regular Form (N = 7,012)	**Customized Form (N = 2,100)**		**Regular Form (N = 51,525)**	**Customized Form (N = 4,686)**
Alternative scoring	1.6	33.4	Alternative scoring	0.7	34.5
Extended-writing options	—	5.9	Extended-writing options	0.2	7.7
Oral administration	—	79.1	Oral administration	0.0	44.8
Signed administration	—	1.3	Signed administration	0.0	1.3
			Audiocassette	0.0	37.7

2.3 TEST ADMINISTRATION TIME

In addition to their demographic information, students were asked to record the times they started and finished the tests. In ELA, students recorded the times for sessions 1 and 2. These times were scanned, and the total testing time was calculated.

In the fall 2004 administration, 83 percent of the students took two hours and thirty minutes or less to finish the mathematics test; in spring 2005, 91 percent finished the mathematics test within that length of time. In session 1 of the fall 2004 ELA administration, 86 percent of the students completed the test within two hours; 81 percent finished within that amount of time in session 2. In session 1 of the spring 2005 administration, 91 percent of the students completed the ELA test within two hours; 89 percent completed the test within that time in session 2. Table 2.6 reports the breakdowns. In the "Invalid" row, table 2.6 also shows that a greater percentage of students left the items blank or recorded invalid numbers in the fall administration than in the spring administration. Total testing times for these students could not be calculated.

TABLE 2.6

Time Taken: Fall 2004 and Spring 2005 (All Attempts)

	Fall 2004			Spring 2005		
	Math (N = 11,373)	ELA (N = 9,112)		Math (N = 58,366)	ELA (N = 56,211)	
	%	%		%	%	
Time Taken		Session 1	Session 2		Session 1	Session 2
15 min	0.7	2.6	1.0	0.3	0.5	0.4
30 min	2.2	11.7	4.2	1.1	3.8	2.1
45 min	5.9	20.5	10.6	3.5	14.4	10.4
1 hr	12.3	21.9	18.0	11.6	24.5	22.8
1 hr 15 min	14.3	12.9	18.3	17.9	21.1	22.6
1 hr 30	15.0	8.3	13.0	19.2	14.1	15.5
1 hr 45	11.9	5.1	9.4	15.2	8.0	9.3
2 hr	10.0	3.5	6.2	11.4	4.7	5.6
2 hr 15 min	6.4	2.2	3.5	6.5	2.2	2.9
2 hr 30 min	4.4	0.9	2.2	3.9	1.1	1.6
2 hr 45 min	2.2	0.7	1.2	1.9	0.5	0.8
3 hr +	3.8	0.9	2.2	2.7	0.9	1.1
Invalid	10.9	9.1	10.3	4.8	4.0	4.9

2.4 STUDENT QUESTIONNAIRES

After the test administration, students were instructed to answer 17 questions for mathematics and 8 questions for ELA on the HSAP student questionnaire. The questionnaire topics encompassed test difficulty, classroom activities, and calculator use (mathematics only).

For mathematics, approximately 85 percent of the students indicated that they did use calculators on the HSAP test (84 percent in fall 2004 and 86 percent in spring 2005), while 9 percent (10 percent in fall 2004 and 7 percent in spring 2005) of the students responded that they did not use calculators. For the math content covered on this test, about 72 percent of the students in spring 2005 and 55 percent of the students in fall 2004 responded that they were familiar with all or most of the test content. About 94 percent of the students reported that they had either more than enough or about the right amount of time for the test. Regarding test difficulty, about 18 percent of the students in spring 2005 and 9 percent of the students in fall 2004 reported that hardly any of the questions were difficult, while about 52 percent of the students felt at least some of the questions were difficult.

For ELA, about 80 percent of the students in spring 2005 and 58 percent of the students in fall 2004 responded that they were familiar with all or most of the content on this test. About 95 percent of the students reported that they had either more than enough or about the right amount of time for the test. For test difficulty, about 25 percent of the students in spring 2005 and 17 percent of the students in fall 2004 reported that hardly any of the questions were difficult. Regarding classroom ELA activities, about 80 percent of the students in spring 2005 and 59

percent of the students in fall 2004 responded that they had written essays or other compositions in their English class (59 percent in fall 2004 and 80 percent in spring 2005). About 51 percent in spring 2005 and 26 percent of the students in fall 2004 responded that they have written poems in their English class. Regarding the types of materials that the students had read in their English class, the category of short stories had the highest number of responses (81 percent in spring 2005 and 66 percent in fall 2004), followed by novels (72 percent in spring 2005 and 48 percent in fall 2004), poems (71 percent in spring 2005 and 43 percent in fall 2004), and plays (62 percent in spring 2005 and 34 percent in fall 2004).

Chapter 3

TEST ADMINISTRATION

3.1 TEST ADMINISTRATION WINDOW

The ELA operational tests for fall 2004 and spring 2005 were conducted in two sessions over two days: October 26–27, 2004, for fall 2004 and April 19–20, 2005, for spring 2004. The mathematics tests were conducted October 28, 2004, for fall and April 21 for spring. The HSAP makeup test window for fall 2004 was from October 29 through November 5, 2004, and for spring, April 22–29, 2005.

The district test coordinators (DTCs) were instructed to administer makeup tests to all eligible students. The administration of one test per day was recommended, but the DTCs were advised that students could take both subjects on one day if necessary.

3.2 TIMING OF THE TESTS

The HSAP tests were not timed; however, students were required to complete each test during a single day (unless a student's IEP or 504 plan specifically stated that he or she needed an administration spanning several days). The following time *estimates* were provided to districts and schools for scheduling purposes only:

ELA, session 1 .....2 hours

ELA, session 2 .....2 hours

Mathematics.....3 hours

In the administration manuals, procedures were outlined for accommodating students who needed time beyond these estimated hours to finish a particular test. Test administrators (TAs) were instructed to give these students as much time as they needed to complete the test, provided that school staff and space were available.

3.3 ADMINISTRATION MANUALS

Working with SDE staff, American Institutes for Research (AIR) staff drafted the administration manuals for the test. SDE staff reviewed and revised the manuals, and the AIR finalized and printed them. Two types of manuals were produced for the HSAP tests: the *HSAP Test Administration Manual* (*TAM*) and the *HSAP District Test Coordinator's Supplement* (SDE 2004b and 2005b). The supplement included only the information that the DTCs needed for the administration of the HSAP tests. The *TAM* contained the information that the school test coordinators (STCs), TAs, and monitors needed to administer the tests to students in their schools.

For both of these administrations, Appendix C in the *TAM* was revised each administration to include a more detailed description of customized materials available for testing and what to order for each group of students requiring specific types of materials. Also updated were the

sections regarding completing student demographic information, returning scorable and nonscorable test materials, and accommodations and modifications tables.

3.4 CUSTOMIZED MATERIALS

Customized versions of the tests were available for ELA and mathematics. Six different customized formats of the HSAP tests were available for these administrations.

- Loose-leaf test booklets, which were printed, single sided, in three-ring binders, allowed individuals to remove the pages so that they could write or type answers to the constructed-response and extended-response items.
- Large-print booklets could be used for students who have difficulty reading text in a standard-size font. The large-print version was printed in a 9 x 12-inch spiral-bound booklet in an 18-point sans serif font.
- Braille booklets were produced for students who typically read classroom materials in braille. The braille versions were spiral bound on 11½- x 11-inch interpoint braille pages.
- A regular-print Form C test booklet was provided in test packets for students or the TAs to use with other customized formats such as the oral script/audiotape; braille, large-print, and loose-leaf versions; and sign language videotapes. These booklets were saddle-stitched and printed in a 12-point font just as the regular, noncustomized test booklets.
- Oral administration scripts and audiotapes were provided for students whose 504 and IEP plans were written to require oral administration of tests. Scripts provided the directions to the TAs regarding the appropriate way to read test questions, passages, and some answer choices to the students. Audiotapes were used for students testing individually or in small-group settings.
- Sign language videotapes were also produced and included the signed test directions, test questions, and some answer choices. The videotapes were produced in three languages: American Sign Language (ASL), Pidgin Signed English (PSE), and Signed Exact English (SEE).

3.5 PRETEST WORKSHOPS AND TRAINING

Pretest workshops were held in Columbia, South Carolina, to train the DTCs and some STCs. The DTCs were invited and could bring up to three additional representatives to the workshop. SDE and AIR staff trained the district staff in attendance. The workshop dates for the fall administration were September 23–24, 2004, and the spring workshops were January 25–27, 2005. A third day was added for the spring workshops to accommodate the schedules of the DTCs.

The AIR was allotted approximately an hour and a half to review the HSAP manuals, security procedures, and any other pertinent information, including an in-depth review of the newly revised instructions for administering tests to students with disabilities. Special focus was given to new procedures or any changes that were implemented for the given administrations.

The *DTC Supplement* and *TAMs* were mailed to the DTCs two weeks before the workshops and were also handed out to the DTCs during the workshop. The DTCs in attendance also received

printed copies of the PowerPoint presentations used during the workshop. In addition, the PowerPoint presentations were posted to the SDE Web site (http://www.ed.sc.gov).

The DTCs were instructed to train all STCs by at least one week prior to testing, and the STCs were instructed to train all TAs and monitors at least one week before test administration.

3.6 MATERIALS SHIPPING AND RETURN

Test materials were shipped to the district offices by Pearson Educational Measurement (PEM), the AIR's subcontractor, and were scheduled to arrive by October 12, 2004, for the fall tests and by April 12, 2005, for the spring tests. The district offices were to distribute these materials to the schools by October 19, 2004, for the fall administration and by April 12, 2005, for the spring administration. Each school's shipment was boxed individually and labeled with the number of boxes shipped for that school. The PEM shipment to each district office also included a 10 percent overage of all test materials—with the exception of customized formats, which were sent only in the quantities ordered. The 10 percent overage was in addition to the 5 percent overage included in school shipments. Overage materials for the districts were to be used by the DTCs to fulfill any additional materials requests from the STCs.

The TAs were instructed to return test materials to their respective STCs immediately after test administration. The STCs redistributed test materials to the TAs who administered makeup tests. Those TAs were instructed to return the makeup materials at the end of the makeup session. The STCs were instructed to return all materials—scorable and nonscorable—to their DTCs within one business day after makeup testing,

With the PEM shipment of overage materials, the DTCs had been sent "district coordinator kits," which included step-by-step directions on how to return scorable and nonscorable materials. These directions listed toll-free phone numbers to call to schedule pickups of returned materials. The DTCs were given specific dates in the manuals for returning materials to PEM. For fall 2004, the DTCs were to return the shipment of scorable materials no later than November 8, 2004; the second shipment was due by November 10, 2004. For spring testing, the first shipment was to consist only of the scorable materials for Graduation Express students and was due to PEM by April 22, 2005. The second shipment of scorable materials was due by May 4, 2005; and the third shipment, which consisted of all nonscorable materials, was due by May 6, 2005.

3.7 TEST SECURITY

The State Board of Education promulgated revised test security regulations (24 S.C. Code Ann. Regs. 43-100) that became effective on June 27, 2003. These regulations were implemented for the first time in the 2004 PACT administration. New test security violations procedures were also developed with the assistance of SLED (State Law Enforcement Division).

Test security prior to, during, and following test administration was regarded as critical. The specific procedures that were followed during the test administration and used in the handling of documentation were those outlined in the *TAM*. Reprinted in this manual are excerpts from Section 59-1-445 of the South Carolina Code of Laws, Section 59-1-447 of the Code, Section 59-30-10(i) of the Code, and State Board of Education Regulation 43-100.

The following guidelines were also included in the *TAM*:

- The STCs were to observe test administration activities and monitor adherence to test security. Examinees were to be made aware that monitoring might occur.
- All secure test materials were required to be kept in a secure, locked location when not in use.
- Before testing, access to secure materials was to be restricted to supervised sessions conducted by the STC. Supervised sessions for coding answer document demographic information could be held the week before testing. Review of test administration directions in oral and signed administration scripts was to be restricted to supervised sessions held after school on the day before each test.
- After testing, access to secure materials was required to be restricted to makeup testing sessions and supervised sessions for completing or editing demographic codes on student answer documents.
- The TAs were to be encouraged to walk around the room during testing to check that students were marking their answers in the correct sections of the answer documents. It was permissible to alert students if they were marking their answers in the wrong sections of the answer documents. However, it was not permissible to stop and read test items or students' responses in their test booklets.

Following the test administration and the return of materials, PEM sent missing materials letters to districts identifying the number of unreturned secure materials and the barcode numbers of each missing document. The districts had two weeks to respond to the letter before PEM and the AIR attempted to contact the DTCs by telephone. Subsequently, the districts either located and returned the materials or sent explanations as to why materials were not found. A toll-free telephone number was provided to answer the DTCs' questions regarding the missing materials; in addition, follow-up procedures were employed until all materials were accounted for.

Secure Materials

It was explained to districts and schools that secure materials included regular-print test booklets and all customized test materials. In addition, reference sheets, scratch paper, and separate pages containing student writing were considered as secure materials and had to be returned with the nonscorable materials after administration of the tests. The DTCs and the STCs were instructed to keep secure materials in locked storage at all times when not in use. These materials were not to be left unattended at any time. Additional security policies requiring secure storage, limited access to items, and secure disposal of documents were explained in the manuals and at the pretest workshops.

Agreements to maintain test security and confidentiality were provided in both manuals, and extras were included in the district and school shipments. The DTCs were instructed to have all persons with access to test materials sign the security agreements if they were not already on file at the district office for the current school year. This necessity was stressed repeatedly in the manuals and during the pretest workshops.

Chapter 4

SCORING

For the fall 2004 and spring 2005 HSAP mathematics and ELA tests, the criteria used to score items were based on the item type. Multiple-choice items were scored using item keys indicating each correct option; constructed-response and extended-response items were scored on the basis of scoring rubrics. For extended-response items, a set of scoring rules was applied in creating final scores. This chapter describes the types of items used on the HSAP assessment, the scoring rules that were applied, and reader reliabilities.

4.1 TYPES OF ITEMS

The 2004–05 HSAP tests included three types of items: multiple choice, constructed response, and extended response.

Multiple Choice

For multiple-choice items, students selected one of four options: A, B, C, or D. Each multiple-choice item was scored as 1 for the correct response and 0 for an incorrect response. Missing responses (i.e., items that a student did not answer at all) and multiple responses were scored as incorrect.

Constructed Response

Constructed-response items were scored using a generic rubric of a 0 to 3 scale. Condition codes of B ("blank") and UR ("unreadable" or "illegible") were used for nonscorable responses. For the purpose of calculating the total score, the condition codes were recoded as 0.

For the purpose of monitoring rater quality, 15 percent of the responses to each constructed-response item by students who had not qualified for Graduation Express were double-read without resolution. The score assigned by the primary reader was taken as the final score for each constructed-response item. A detailed scoring rubric providing descriptions of the various score points was used in the scoring process.

For the Graduation Express students, all constructed-response item responses were read by two raters. The final score was determined on the basis of the following rules:

- If the first reader's score was equal to the second reader's score, the reported score was the first reader's score.
- If the first reader's score was different from the second reader's score, a resolution was required.
- If the third reader's score agreed exactly with the first or the second reader's score, the third reader's score was the resolution score.
- If the third reader's score was different from the first or the second reader's score, the reported score was the adjudication score.

Extended Response

An extended-response writing item was administered at the beginning of session 1 of the ELA test and was scored under four domains: content and development, organization, voice, and conventions. Score ranges for these domains are 1–4 for content and development, 1–4 for organization, 1–3 for voice, and 1–4 for conventions, for a total possible score of 15 points. Each extended-response item was independently read by two raters, for a total possible composite score of 30 points. In addition to the double scoring, about 8 percent of the papers were back-read by chief readers.

For the nonscorable responses, condition codes of B ("blank"), OT ("off topic"), IS ("insufficient" response), and UR ("unreadable" or "illegible response") were assigned. For scoring purposes, the condition codes were recoded as 0. The algorithm for scoring extended-response writing is presented in table 4.1 for scorable responses (e.g., 1–4 or 1–3 for domain scores). When a paper received a condition code, the paper was pulled and scored by supervisors. The scoring rules for these papers are presented in table 4.2. As with the constructed-response items, the extended-response items were also scored with a detailed rubric that was generic across all extended-response items.

For the Graduation Express students, each extended-response item was independently scored by two raters. The two scores were processed according to the scoring algorithms in tables 4.1 and 4.2 to produce a final score.

TABLE 4.1

Extended-Response Writing Scoring Algorithm for Papers with Scorable Responses

	First Score (R1)	Second Score (R2)	Action	Back Reading (BR)	Resolution Score (RS) [Third Score]	Final Score (F)
1	R1 = 1–4	R2 = R1	none	NA		F = R1 + R2
2	R1 = 1–4	R2 = 1–4 *and* is adjacent to R1	none	NA		F = R1 + R2
3	R1 = 1–4	R2 = 1–4 *and* is nonadjacent to R1	resolution required	NA	RS = R1	F = RS + R1
4	R1 = 1–4	R2=1–4 *and* is nonadjacent to R1	resolution required	NA	RS = R2	F = RS + R2
5	R1 = 1–4	R2 = 1–4 *and* is nonadjacent to R1	resolution required	NA	RS is adjacent to R1 and R2	F = RS + RS
6	R1 = 1–4	R2 = 1–4 *and* is nonadjacent to R1	resolution required	NA	RS is adjacent to R1 or R2 but not both	F = RS + R1 if R1 is closer to RS than R2 F = RS + R2 if R2 is closer to RS than R1
7	R1 = 1–4	R2 = R1		BR = R1 = R2		F = BR + R1
8	R1 = 1–4	R2 = R1		BR is adjacent to R1 and R2		F = BR + R1

TABLE 4.1

Extended-Response Writing Scoring Algorithm for Papers with Scorable Responses

	First Score (R1)	Second Score (R2)	Action	Back Reading (BR)	Resolution Score (RS) [Third Score]	Final Score (F)
9	R1 = 1–4	R2 = R1		BR is nonadjacent to R1 and R2		F = BR + BR
10	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1		BR = R1 and adjacent to R2		F = BR + R1
11	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1		BR = R2 and adjacent to R1		F = BR + R2
12	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1		BR is adjacent to R1 and discrepant to R2		F = BR + R1
13	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1		BR is adjacent to R2 and discrepant to R1		F = BR + R2
14	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1		BR is nonadjacent to R1 and R2		F = BR + BR

TABLE 4.2

Extended-Response Writing Scoring Algorithm for Papers with Condition Codes

Rule	Supervisor First Score (S1)	Supervisor Second Score (S2)	Action	BR	Supervisor Resolution Score (S3)	Final Score (F)
1	S1 = condition code	S2 = S1	none	NA		F = S1
2	S1 = 1–4	S2 = condition code	resolution required	NA	S3 = 1–4	F = S3 + S1
3	S1 = condition code	S2 = 1–4	resolution required	NA	S3 = 1–4	F = S3 + S2
4	S1 = 1–4	S2 = condition code	resolution required	NA	S3 = condition code	F = S3
5	S1 = condition code	S2 = condition code but not equal to S1	resolution required	NA	S3 = condition code	F = S3
6	S1 = condition code	S2 = condition code but not equal to S1	resolution required	NA	S3 = 1–4	F = S3 + S3

4.2 TEST SPECIFICATIONS

The 2004–05 HSAP test specifications for mathematics and ELA are shown in tables 4.3 and 4.4. As noted previously, the 2004–05 HSAP assessments included multiple-choice, constructed-response, and extended-response items. The integrated-response items are 3-point constructed-response items that integrate content standards and process standards; they require students to use the process skills of problem solving, communication, representations, and connections to apply a solution strategy and then to communicate and represent the result.

TABLE 4.3

Mathematics: Distribution of Score Point Values by Reporting Category

Fall 2004 and Spring 2005	Algebra	Data Analysis and Probability	Measurement and Geometry	Number and Operations	Integrated Responses
Percentage	27%	11%	27%	23%	13%
Multiple-choice points	19	8	19	6	—
Constructed-response points	—	—	—	—	9

TABLE 4.4

ELA: Distribution of Score Point Values by Reporting Category

	Reading Process and Comprehension	Analysis of Texts	Word Study and Analysis	Research	Writing
Fall 2004					
Percentage	24%	16%	13%	7%	40%
Multiple-choice points	20	15	9	7	8
Constructed-response points	3	—	3	—	—
Extended-response points	—	—	—	—	30
Spring 2005					
Percentage	24%	20%	8%	8%	40%
Multiple-choice points	20	16	8	8	8
Constructed-response points	3	3	—	—	—
Extended-response points	—	—	—	—	30

4.3 SCORING PROCESS

PEM scored all items: multiple-choice items were scored by PEM's electronic scanning system; constructed-response (CR) and extended-response (ER) items were scored by trained personnel using the ePEN (Electronic Performance Evaluation Network) system. There were three scoring sites: ELA was scored in Mesa, Arizona, for the fall 2004 administration and in Houston, Texas, for the spring 2005 administration and were supplemented by scorers in Mesa for the spring

2005 administration. Mathematics was scored in Lansing, Michigan, for fall 2004 and spring 2005.

Prior to actual scoring of the constructed-response and extended-response items, rangefinding meetings were held in Columbia, South Carolina, September 20–22, 2004, for the fall 2004 administration and February 2–4, 2005, for the spring 2005 administration. The purposes of the range-finding meetings were twofold: to identify sets of papers that were representative of the various performance levels defined by the rubric and to arrive at consensus scores on large sets of papers for use in training raters. Three range-finding committees—one each for reading, writing, and mathematics—were convened. The committees were composed of educators from South Carolina and were selected by the SDE. Each committee reviewed several items. That is, each committee reviewed multiple papers (students' responses written to a specific item) for multiple items.

AIR and SDE staff were on-site during the first week of rater training (scorers received on-line training via the ePEN system) and live scoring and monitored the scoring process until scoring was complete. Throughout the scoring process, PEM staff posted the performance of each reader (reader-reliability statistics) once a day on PEM's SchoolHouse Web site for AIR and SDE staff to review.

Before start scoring of the live CR and ER items, readers had to pass two of three qualifying sets. Each qualifying set consists of 20 papers. The qualification requirement is as follows:

- ELA ER: 70 percent exact and 80 percent adjacent on 2 of 3 sets with 20 papers in each set
- ELA CR: 75 percent exact and 90 percent adjacent on 2 of 3 sets with 20 papers in each set
- Math CR: 80 percent exact and 90 percent adjacent on 2 of 3 sets with 20 papers in each set

Throughout scoring, readers' performances were monitored through the use of validity papers, which are prescored responses distributed to readers throughout scoring to ensure that the readers, as well as scoring supervisors, do not drift from the scoring rubric. "True scores" for these papers were assigned by scoring leaders and then stored in the ePEN system. Reader agreement was checked on a regular basis—every twenty papers for the extended-response item and every sixty papers for CR items. This quality check was "blind" in that readers did not know they were scoring a validity paper.

4.4 READER RELIABILITY

In the scoring of constructed-response and extended-response items, 15 percent of the papers for CR items and 100 percent of the papers for ER items were independently scored by two readers. The percentages of reader consistency on the papers that were double-scored are reported in tables 4.5 and 4.6.

The reported reader-reliability indexes are rates of perfect agreement and rates of perfect and adjacent agreement. The term *perfect agreement* indicates that the two readers assigned the same score to the same written response. The term *adjacent agreement* indicates that the two readers differed by 1 point when evaluating the same response.

TABLE 4.5

Mathematics: Reader Reliabilities for Scoring Constructed-Response Items

Items	Number	Perfect Agreement	Perfect and Adjacent Agreement
Fall 2004			
CR1	1,710	82.5%	99.3%
CR2	1,715	90.1%	99.1%
CR3	1,700	93.6%	99.5%
Spring 2005			
CR1	7,744	90.8%	99.5%
CR2	7,741	86.5%	99.1%
CR3	7,772	89.0%	98.7%

TABLE 4.6

ELA: Reader Reliabilities for Scoring Constructed-Response and Extended-Response Items

Items	Number	Perfect Agreement	Perfect and Adjacent Agreement
Fall 2004			
CR1	1,305	68.6%	97.9%
CR2	1,367	66.7%	97.1%
ER content and development	8,839	64.8%	98.8%
ER organization	8,839	62.6%	98.1%
ER voice	8,839	66.8%	99.0%
ER conventions	8,839	54.9%	93.9%
Spring 2005			
CR1	7,625	68.7%	98.2%
CR2	7,774	71.7%	98.8%
ER content and development	51,459	73.5%	99.5%
ER organization	51,459	69.6%	98.5%
ER voice	51,459	73.3%	99.5%
ER conventions	51,459	75.7%	98.0%

4.5 TESTED/NOT TESTED FLAG

A student was considered "tested" in mathematics if he or she answered at least one question. The question could have been a multiple-choice or constructed-response item. A student was considered "tested" in ELA if he or she answered at least one question on either of the two days of testing. The one question could have been a multiple-choice item, constructed-response item, or extended-response writing item.

Chapter 5

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. The term *item discrimination* refers to a correlation between the student's item score and the student's total score. For the discrimination index of a particular item, point-biserial correlations were produced. In the calculation of the point-biserial correlation for a particular item, that item was excluded from the total score.

A "not-reached" (NR) item was any one to which a student did not respond after the last item that he or she attempted in a session. In other words, an item was not reached if the student did not respond to it or to any other item after it. An "omit" was any nonresponse item appearing between items with responses.

In recoding missing data for item analysis, all omitted and NR items were recoded as incorrect, with a zero score. After holding discussions, SDE and AIR staff decided to exclude from the CTT item analyses and item calibrations those students who had used customized materials and those who had received the alternative scoring rubric modification.

5.1 ITEM NONRESPONSE RATES

Although the HSAP tests were not timed, students were required to finish each test session during one school day, unless they had an IEP that allowed for accommodations in administration. The TAs were instructed that the expected test duration for each ELA session would be about two hours and that the mathematics test could be expected to run approximately three hours.

The percentage of students who responded to the last two items on a given test form was computed. Table 5.1 presents the average of these percentages across the different forms for each subject. The percentages listed in the "Last Item" column of the table represent those students who responded to the last item—constructed-response (CR) item 3 for mathematics, a multiple-choice (MC) item in both sessions 1 and 2 for ELA. The percentages in the adjacent column include students who omitted the last item on the test but answered the second-to-last item—CR item 2 for mathematics, item 14 in fall and item 21 in spring in session 1 and item 59 in fall, and item 73 in spring in session 2 for ELA. Item nonresponse rates were computed for each ELA session separately. Students tend to leave CR items blank more often than they leave MC items blank, especially when the CR items appear at the end of the test.

TABLE 5.1

Percentage of Students Responding to Last and Second-to-Last Items

Fall 2004			Spring 2005		
Subject	**Last Item**	**Second-to-Last Item**	**Subject**	**Last Item**	**Second-to-Last Item**
Mathematics	86.0 (CR)	87.4 (CR)	Mathematics	89.1 (CR)	95.9 (CR)
ELA session 1	96.8 (MC)	97.3 (MC)	ELA session 1	99.0 (MC)	99.1 (MC)
ELA session 2	98.0 (MC)	97.2 (MC)	ELA session 2	99.2 (MC)	99.1 (MC)

5.2 CLASSICAL ITEM STATISTICS

Table 5.2 provides a summary of item *p*-values and item discriminations by item types and content areas for the mathematics operational items and a summary of item *p*-values and item discriminations by item types and content areas for the ELA operational and embedded field-test items. For CR and ER items, the *p*-value was computed as the ratio of the item mean to the item's maximum possible score (MPS). For the discrimination index, point-biserial correlations were computed between the item and the total raw score as the criterion. In the computing of the point-biserial correlation, the item was excluded in the total raw score.

TABLE 5.2

Summary of Classical Item Statistics for Mathematics and ELA

Item Type/ Content Area	Number of Items	*p*-value	Point–Biserial Correlation	Number of Items	*p*-value	Point–Biserial Correlation
		Fall 2004			Spring 2005	
Mathematics						
Multiple-choice	62	0.50	0.30	62	0.67	0.42
Constructed-response	3	0.26	0.62	3	0.56	0.71
Number and Operations	16	0.57	0.31	16	0.73	0.45
Algebra	19	0.54	0.29	19	0.68	0.40
Measurement and Geometry	19	0.41	0.31	19	0.60	0.41
Data Analysis and Probability	8	0.47	0.29	8	0.69	0.44
ELA						
Multiple-choice	60	0.56	0.33	60	0.72	0.35
Constructed-response	2	0.48	0.60	2	0.44	0.54
Extended-response	1	0.71	0.74	8	0.90	0.63
Reading Process and Comprehension	21	0.57	0.35	21	0.78	0.37
Analysis of Texts	16	0.56	0.37	17	0.62	0.35
Word Study and Analysis	10	0.55	0.33	8	0.78	0.37
Research	7	0.50	0.30	8	0.72	0.33
Writing	9	0.63	0.53	16	0.78	0.50
Field-test items				80*	0.72	0.38

*A total of 8 sets of 10 items (5 items in each session) were appended to the ELA base form, resulting in a total of 80 field-test items across 8 ELA forms.

Chapter 6

ITEM CALIBRATION AND SCALING

6.1 METHODOLOGY AND SOFTWARE

The Rasch model was used in the item calibrations of the HSAP items. The one-parameter Rasch model (Rasch 1980; Wright and Stone 1979) was used to calibrate multiple-choice items. Constructed-response and extended-response items were calibrated with the Rasch partial credit model (Masters 1982). Calibrating mixed item types from different assessment modes (i.e., dichotomously and polytomously scored items) requires the use of a polytomous model, which allows the number of score categories (typically score points on a scoring rubric) to vary across assessment modes. The Rasch partial credit model (Wright and Masters 1982) can accommodate the mixing of dichotomous and polytomous items.

The Rasch partial credit model is widely used for high school graduation exams, particularly those with high stakes for students and educators. The AIR used a one-to-one translation from the number of correct responses to the scale score in the Rasch model. Maintaining a correspondence between the raw number correct score and the scale score, while simultaneously equating multiple test forms, posed a challenge that was best met by using the one-parameter Rasch dichotomous model and the Rasch partial credit model (Wright and Masters 1982).

The WINSTEPS software program (Linacre and Wright 2003) was used in the item calibration. WINSTEPS uses the joint maximum-likelihood estimation (JMLE) approach, which estimates the item and person parameters simultaneously. Although this estimation method is subject to small statistical biases, which increase as the length of the scale decreases, these biases were corrected through the use of the WINSTEPS feature STBIAS=Y.

6.2 ITEM CALIBRATION

For mathematics, the equated operational test forms were constructed from the precalibrated item pool based on the spring 2003 census field-test items; therefore, the raw-score-to-scale-score conversion table for the fall 2004 and spring 2005 operational forms were created before the test was administered.

For ELA, although the spring 2003 field-test forms covered all standards specified in the ELA test specification, the number of items for a few academic standards needed to be increased. In order to replenish the ELA precalibrated item pool for these standards, the SDE and the AIR decided to embed field-test items in spring 2004 and 2005 HSAP operational administrations. In the spring 2005 HSAP administration, 67 items (64 MC, 2 CR, and 1 ER) were common on all ELA forms. The 64 multiple-choice items included 57 operational items and 7 embedded field-test items. In each form, 10 unique field-test items (5 items at the end of each session) were added, resulting in a total of 80 unique field-test items for future use.

In ELA, the field-test items (including all embedded and added field-test items) were placed on the item bank scale. The operational item parameters were anchored at the bank difficulty values; therefore, the field-test item difficulties were mapped onto the bank metric in the concurrent calibration.

6.3 COMPOSITION OF THE CALIBRATION SAMPLE

A subset of the embedded field-test items was expected to be used as operational items in order to fulfill test blueprint requirements. Early return samples were identified so that parameter estimation for the field-test items could begin as soon as possible after test administration and not jeopardize the score reporting schedule.

The samples were preselected based on spring 2004 HSAP results (i.e., ELA mean scale score) which was used as a sampling stratum. All regular schools were stratified into deciles on the basis of their spring 2004 ELA performance. Four schools were randomly selected from each decile for a total of 40 schools from 28 districts.

The eight ELA forms were spiraled within classrooms to reduce the "cluster effect." (Because students within a school or classroom tend to be more similar to each other than to students statewide, they bring less unique information to a sample—an effect called "clustering.") Clustering can dramatically reduce the amount of information in a sample and increase the error variance. When data are clustered, the effective sample size is used to indicate the number of independent pieces of information available from the clustered sample. This is essentially equal to the size of a simple random sample that would yield the same precision as the clustered sample. The degree of dependence within a clustered sample is indicated by the intraclass correlation coefficient. The higher the intraclass correlation, the less independence there is within the clustered sample. A simple random sample would have an intraclass correlation equal to zero.

For the ELA calibration sample, a sample size of 40 schools (1,155 students per form) is equivalent to an effective sample size of 387 students per form from a simple random sampling of students. The effective sample size is based on an average school size of 231 students who took the ELA test, an average class size of 20, a class intraclass correlation (ρ_c = .323), and a school intraclass correlation (ρ_s = .057).

Because 19 small districts of the 28 districts returned their test materials early for all their schools, not just their sampled schools, the obtained sample included 77 schools from 28 districts (1,600 students per form). For the item calibration analysis, students who took a regular form, students who attempted the test for the first time, and students who were in grades nine and ten were included.

6.4 SCALING

Based on the precalibrated item pool, Rasch-ability-score-to-scale-score conversion tables were generated for each subject. These 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 the transformation of Rasch-ability-score-to-scale-score, the following steps were taken in generating scale scores:

Step 1: Linear transformation of Rasch-ability-score-to-scale-score, fixing the passing scale score (Level 2) at 200 with a standard deviation of 25,

$$SS = SS_C + B\left[\frac{\hat{\theta} - \theta_C}{\sigma_{\hat{\theta}}}\right] = 200 + 25\left[\frac{\hat{\theta} - \theta_C}{\sigma_{\hat{\theta}}}\right],$$

where the passing ability scores (θ_C) are –0.224 for mathematics and 0.015 for ELA and the standard deviations of *theta* ($\sigma_{\hat{\theta}}$) are 1.102 for mathematics and 1.046 for ELA.

Step 2: The decimals in the scale score were truncated to avoid the same scale score for two different raw scores.

Step 3: Scale scores less than 100 and greater than 320 were fixed at 100 and 320, respectively.

6.5 DEFINITION OF SCOREABILITY

A student was considered "tested" if the student answered at least one question in the test booklet. All tested students' item responses were scored. All omits and not-reached items were recoded as incorrect and given a zero score.

6.6 REPORTING OF ZERO AND PERFECT SCORE

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 the extreme values. That is, a fractional score point value was subtracted from perfect scores, and was added to zero scores.

6.7 POLICY DEFINITION OF ACHIEVEMENT LEVELS

After the spring 2003 HSAP census field test, the AIR, in collaboration with its partner Insite, Inc., conducted standard-setting workshops for the HSAP mathematics and ELA examinations on July 21–25, 2003. In each subject, the workshop participants recommended three achievement-level cut scores: Level 2, Level 3, and Level 4. Level 2 was the cut required for student graduation purposes, and Levels 3 and 4 described students for AYP (adequate yearly progress) purposes. Achievement-level descriptions are provided below in tables 6.1 and 6.2. The AIR outlined the details of the standard-setting process in its 2004 report to the SDE, "South Carolina High School Assessment Program English Language Arts and Mathematics Standard Setting Technical Report."

TABLE 6.1

Description of Achievement Levels for the HSAP Mathematics Test

Level	Description
4	The Level 4 student • has demonstrated an exceptional command of skills and knowledge required of high school students in South Carolina • analyzes, evaluates, and/or synthesizes mathematical concepts and procedures and solves problems using advanced arithmetic, algebraic, and measurement/geometric concepts and relationships • analyzes data representations and applies probability concepts • supports answers with mathematical work and/or explanations that thoroughly communicate mathematical reasoning • has met the exit examination requirement for a South Carolina high school diploma
3	The Level 3 student • has demonstrated proficiency in skills and knowledge required of high school students in South Carolina • applies mathematical concepts and procedures and solves problems using arithmetic, algebraic, and measurement/geometric concepts and relationships • interprets data representations and demonstrates a knowledge of probability concepts • supports answers with mathematical work and/or explanations that clearly communicate mathematical reasoning • has met the exit examination requirement for a South Carolina high school diploma
2	The Level 2 student • has demonstrated competence in skills and knowledge required of high school students in South Carolina • demonstrates an acceptable knowledge of fundamental mathematical concepts and procedures and solves problems using essential arithmetic, algebraic, and measurement/geometric concepts and relationships • demonstrates a knowledge of basic data representations and probability concepts • supports answers with mathematical work and/or explanations that adequately communicate mathematical reasoning • has met the exit examination requirement for a South Carolina high school diploma
1	The Level 1 student • has not demonstrated competence in the skills and knowledge required of high school students in South Carolina • demonstrates a limited understanding of mathematical concepts • is able to use arithmetic, algebraic, and measurement/geometric concepts and relationships • demonstrates a knowledge of simple data representations and probability concepts • supports answers with mathematical work and/or explanations that minimally communicate mathematical reasoning • has not met the exit examination requirement for a South Carolina high school diploma

TABLE 6.2

Description of Achievement Levels for the HSAP ELA Test

Level	Description
4	The Level 4 student • has demonstrated an exceptional command of skills and knowledge required of high school students in South Carolina • demonstrates comprehension of complex ideas and connects those ideas within a text, across texts, and beyond the text • displays exceptional writing skills by engaging the reader, effectively developing and organizing ideas, and using relevant supporting details, vivid language, and Standard American English • has met the exit examination requirement for a South Carolina high school diploma
3	The Level 3 student • has demonstrated proficiency in skills and knowledge required of high school students in South Carolina • demonstrates comprehension of complex ideas and connects those ideas within a text and across texts • displays effective writing skills by sustaining the reader's interest, clearly developing and organizing ideas, and using relevant supporting details and Standard American English • has met the exit examination requirement for a South Carolina high school diploma
2	The Level 2 student • has demonstrated competence in skills and knowledge required of high school students in South Carolina • demonstrates comprehension of essential ideas and shows some logical connections of those ideas within a text • displays acceptable writing skills by showing some awareness of audience, developing and organizing ideas, and using relevant supporting details and Standard American English • has met the exit examination requirement for a South Carolina high school diploma
1	The Level 1 student • has not demonstrated competence in skills and knowledge required of high school students in South Carolina • demonstrates limited comprehension of ideas and tenuous connections of those ideas within a text • displays limited writing skills, which may include little awareness of audience and purpose, partial development and organization of ideas, and deviations from Standard American English • has not met the exit examination requirement for a South Carolina high school diploma

6.8 CUT SCORE FOR ACHIEVEMENT LEVELS

The cut scores for total scores for operational HSAP test forms are presented in table 6.3.

TABLE 6.3

Cut Scores in Rasch Ability Scale and Scale Score for Total Score

	Level 2	Level 3	Level 4
Mathematics			
Rasch Ability	-0.224	0.658	1.584
Scale Score	200	220	241
ELA			
Rasch Ability	0.015	0.978	1.731
Scale Score	200	223	241

These cut scores were derived from the HSAP standard-setting study and do not vary across test forms.

6.9 CONTENT-AREA INFORMATION

In addition to total scores, information was reported for four content areas in mathematics and five content areas in ELA. For each content area, the following steps were taken:

Step 1: A raw-score-to-Rasch-ability-score conversion table was generated for each content area. The empirical Level 2 cut score (i.e., the raw score with the smallest Rasch ability value equal to or greater than the Level 2 Rasch ability cut score for the total test) was located on each content-area scale.

Step 2: A 68 percent confidence interval of the cut score (θ_c) was computed as cut score (θ_c) $\pm$ 1 SE(θ_c). The scores were grouped into one of three classifications as follows:

Adequate: if $\theta \geq \theta_c + 1$ SE

May need improvement: if $\theta_c - 1$ SE $\leq \theta < \theta_c + 1$ SE

Needs improvement: if $\theta < \theta_c - 1$ SE

The empirical Rasch-ability-score-to-content-area cut scores used for the three classifications for each content area are provided in table 6.4.

TABLE 6.4

Cut Scores on the Rasch Ability Scale, Associated Standard Errors, and Confidence Intervals for Content-Area Classifications

Content Area	Rasch Ability (θ)	SE(θ)	68% Confidence Interval	
			θ – 1SE	θ + 1SE
Mathematics				
Fall 2004				
Number and Operations	0.012	0.557	–0.545	0.569
Algebra	0.004	0.500	–0.496	0.504
Measurement and Geometry	–0.058	0.475	–0.533	0.417
Data Analysis and Probability	–0.066	0.746	–0.812	0.680
Spring 2005				
Number and Operations	0.045	0.555	–0.510	0.600
Algebra	–0.192	0.487	–0.679	0.295
Measurement and Geometry	–0.042	0.487	–0.529	0.445
Data Analysis and Probability	–0.166	0.728	–0.894	0.562
ELA				
Fall 2004				
Reading Process and Comprehension	0.101	0.444	–0.343	0.545
Analysis of Texts	0.098	0.535	–0.437	0.633
Word Study and Analysis	0.228	0.616	–0.388	0.844
Research	0.525	0.771	–0.246	1.296
Writing	0.142	0.414	–0.272	0.556
Spring 2005				
Reading Process and Comprehension	0.167	0.442	–0.275	0.609
Analysis of Texts	0.168	0.488	–0.320	0.656
Word Study and Analysis	0.190	0.752	–0.562	0.942
Research	0.538	0.756	–0.218	1.294
Writing	0.159	0.406	–0.247	0.565

6.10 PERCENTAGE OF STUDENTS IN EACH ACHIEVEMENT LEVEL

Tables 6.5 through 6.8, below, present student performance on the fall 2004 and spring 2005 HSAP operational tests for mathematics and ELA. Percentages of students in the four achievement levels are reported for overall and subgroups. Subgroups include the reporting categories of gender, ethnicity, language fluency (i.e., LEP—limited English proficiency), lunch program participation, migrant status, and disability. The summary includes all students who were tested but excludes students in adult education and district-approved homeschools. Tables 6.9 though 6.12 provide the information for content areas. The information is summarized for Level 1 and at or above Level 2 for all students by gender and by ethnic group. Of those students who took both the mathematics and ELA tests for the first time, 74 percent in spring 2005 passed both tests. In fall 2004, of those students who took both tests one or multiple times, 23 percent passed both tests.

TABLE 6.5

Fall 2004 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts)

Subgroup	Achievement Levels				L2+*	L3+**	N
	Level 1	Level 2	Level 3	Level 4			
Overall	69.3	22.9	4.8	3.0	30.7	7.8	11,373
Gender							
Female	67.0	24.9	5.0	3.2	33.0	8.1	5,099
Male	71.6	21.1	4.6	2.8	28.4	7.3	5,977
Invalid	60.9	27.3	7.4	4.4	39.1	11.8	297
Ethnicity							
African American	77.1	20.8	1.7	0.4	22.9	2.1	6,969
Asian/Pacific Islander	38.6	20.0	18.6	22.9	61.4	41.4	70
Hispanic	65.0	24.6	7.8	2.7	35.0	10.5	334
American Indian	55.0	45.0	0.0	0.0	45.0	0.0	20
White	56.7	26.1	10.0	7.3	43.3	17.2	3,566
Other	46.4	31.2	12.3	10.1	53.6	22.5	138
Unknown	58.7	28.6	7.6	5.1	41.3	12.7	276
Language							
English speaker	69.3	22.9	4.8	3.0	30.7	7.8	11,188
Full LEP	67.2	20.2	9.2	3.4	32.8	12.6	119
LEP mainstream	62.9	25.7	8.6	2.9	37.1	11.4	35
Waiver	60.0	40.0	0.0	0.0	40.0	0.0	10
Exited	61.9	28.6	4.8	4.8	38.1	9.5	21
Lunch Program							
No free/reduced lunch	66.4	24.2	5.6	3.8	33.6	9.5	7,751
Free lunch	76.4	19.7	2.9	1.0	23.6	3.9	3,239
Reduced lunch	66.8	25.6	5.2	2.3	33.2	7.6	383
IEP							
Yes	89.3	9.3	1.3	0.2	10.7	1.4	2,512
No	63.6	26.8	5.9	3.8	36.4	9.6	8,852
Unknown	77.8	22.2	0.0	0.0	22.2	0.0	9
Migrant							
Yes	70.6	17.6	5.9	5.9	29.4	11.8	17
No	69.2	22.9	4.8	3.0	30.8	7.8	11,356

***** indicates the percentage of students at or above Level 2
** indicates the percentage of students at or above Level 3

TABLE 6.6

Spring 2005 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (First Attempt)

Subgroup	Achievement Levels				L2+*	L3+**	N
	Level 1	Level 2	Level 3	Level 4			
Overall	24.6	29.1	25.0	21.3	75.4	46.3	51,270
Gender							
Female	22.7	31.6	25.5	20.1	77.3	45.6	25,570
Male	26.0	26.6	24.6	22.8	74.0	47.4	25,344
Invalid	51.1	26.4	15.7	6.7	48.9	22.5	356
Ethnicity							
African American	39.8	35.1	18.3	6.8	60.2	25.1	20,226
Asian/Pacific Islander	10.3	19.5	25.2	45.0	89.7	70.2	584
Hispanic	31.0	33.0	22.7	13.3	69.0	36.0	1,260
American Indian	21.3	35.1	28.7	14.9	78.7	43.6	94
White	13.3	24.8	30.0	32.0	86.7	61.9	28,254
Other	26.4	32.2	26.0	15.3	73.6	41.3	450
Unknown	48.8	27.9	15.9	7.5	51.2	23.4	402
Language							
English speaker	24.3	29.1	25.1	21.5	75.7	46.6	50,409
Full LEP	52.1	30.2	12.7	5.0	47.9	17.7	424
LEP mainstream	25.0	43.3	20.7	11.0	75.0	31.7	164
Waiver	19.6	26.1	32.6	21.7	80.4	54.3	46
Exited	24.8	24.8	25.7	24.8	75.2	50.4	226
Unknown	0.0	0.0	0.0	100.0	100.0	100.0	1
Lunch Program							
No free/reduced lunch	15.3	25.6	28.6	30.5	84.7	59.1	29,942
Free lunch	39.8	34.0	18.8	7.4	60.2	26.2	17,997
Reduced lunch	25.6	33.7	26.0	14.7	74.4	40.7	3,331
IEP							
Yes	68.9	20.6	8.0	2.6	31.1	10.5	6,258
No	18.4	30.3	27.4	24.0	81.6	51.3	44,976
Unknown	66.7	27.8	2.8	2.8	33.3	5.6	36
Migrant							
Yes	41.9	20.9	25.6	11.6	58.1	37.2	43
No	24.5	29.1	25.0	21.4	75.5	46.3	51,227

* indicates the percentage of students at or above Level 2
** indicates the percentage of students at or above Level 3

TABLE 6.7

Fall 2004 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts)

Subgroup	Achievement Levels				L2+*	L3+**	N
	Level 1	Level 2	Level 3	Level 4			
Overall	57.0	29.8	7.9	5.3	43.0	13.2	9,112
Gender							
Female	49.2	34.0	9.4	7.5	50.8	16.8	3,512
Male	62.3	27.2	6.7	3.8	37.7	10.5	5,335
Invalid	54.3	26.0	14.0	5.7	45.7	19.6	265
Ethnicity							
African American	64.1	31.1	3.9	1.0	35.9	4.9	5,430
Asian/Pacific Islander	48.4	28.6	15.4	7.7	51.6	23.1	91
Hispanic	61.0	27.5	7.0	4.4	39.0	11.4	385
American Indian	68.8	12.5	18.8	0.0	31.3	18.8	16
White	43.9	28.5	14.6	13.0	56.1	27.6	2,838
Other	40.7	29.2	19.5	10.6	59.3	30.1	113
Unknown	56.1	21.8	13.8	8.4	43.9	22.2	239
Language							
English speaker	56.7	29.8	8.1	5.4	43.3	13.5	8,871
Full LEP	73.5	24.7	1.8	0.0	26.5	1.8	166
LEP mainstream	59.5	35.7	4.8	0.0	40.5	4.8	42
Waiver	45.5	45.5	9.1	0.0	54.5	9.1	11
Exited	54.5	31.8	9.1	4.5	45.5	13.6	22
Lunch Program							
No free/reduced lunch	54.0	29.7	9.2	7.1	46.0	16.3	5,982
Free lunch	63.9	29.6	5.0	1.5	36.1	6.6	2,819
Reduced lunch	53.7	32.8	9.3	4.2	46.3	13.5	311
IEP							
Yes	79.8	17.7	2.0	0.5	20.2	2.4	2,334
No	49.2	33.9	10.0	6.9	50.8	16.9	6,773
Unknown	40.0	60.0	0.0	0.0	60.0	0.0	5
Migrant							
Yes	75.0	10.0	10.0	5.0	25.0	15.0	20
No	57.0	29.8	7.9	5.3	43.0	13.2	9,092

***** indicates the percentage of students at or above Level 2
** indicates the percentage of students at or above Level 3

TABLE 6.8

Spring 2005 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (First Attempt)

Subgroup	Achievement Levels				L2+*	L3+**	N
	Level 1	Level 2	Level 3	Level 4			
Overall	14.5	30.7	29.2	25.6	85.5	54.9	51,459
Gender							
Female	10.4	30.8	30.4	28.4	89.6	58.8	25,629
Male	18.3	30.5	28.1	23.0	81.7	51.2	25,515
Invalid	34.3	34.6	21.6	9.5	65.7	31.1	315
Ethnicity							
African American	23.6	42.6	24.7	9.2	76.4	33.9	20,348
Asian/Pacific Islander	8.9	23.5	28.1	39.6	91.1	67.6	584
Hispanic	25.2	31.6	27.1	16.1	74.8	43.2	1,274
American Indian	9.5	31.6	30.5	28.4	90.5	58.9	95
White	7.3	22.2	32.7	37.8	92.7	70.5	28,351
Other	18.0	31.5	26.1	24.3	82.0	50.5	444
Unknown	31.7	34.7	24.2	9.4	68.3	33.6	363
Language							
English speaker	14.1	30.6	29.4	25.9	85.9	55.3	50,591
Full LEP	57.1	32.5	8.8	1.6	42.9	10.4	431
LEP mainstream	27.9	38.8	23.0	10.3	72.1	33.3	165
Waiver	23.4	34.0	36.2	6.4	76.6	42.6	47
Exited	10.3	30.4	32.1	27.2	89.7	59.4	224
Unknown	0.0	0.0	0.0	100.0	100.0	100.0	1
Lunch Program							
No free/reduced lunch	8.1	23.3	32.3	36.4	91.9	68.7	30,004
Free lunch	25.0	41.8	23.9	9.3	75.0	33.2	18,114
Reduced lunch	14.4	37.1	30.4	18.1	85.6	48.5	3,341
IEP							
Yes	56.0	30.5	10.5	3.0	44.0	13.5	6,309
No	8.6	30.7	31.8	28.8	91.4	60.7	45,114
Unknown	50.0	30.6	11.1	8.3	50.0	19.4	36
Migrant							
Yes	22.5	50.0	17.5	10.0	77.5	27.5	40
No	14.5	30.7	29.2	25.7	85.5	54.9	51,419

***** indicates the percentage of students at or above Level 2
** indicates the percentage of students at or above Level 3

TABLE 6.9

Fall 2004 HSAP Mathematics Operational Test: Content-Area Information (All Attempts)

Subgroup	Level 1				Level 2 and Above			
	Needs Improvement	*May need improvement*	*Adequate*	N1*	*Needs Improvement*	*May need improvement*	*Adequate*	N2**
Number and Operations								
All students	82.2%	16.7%	1.1%	7,876	15.6%	46.1%	38.3%	3,497
Females	81.6%	17.6%	0.8%	3,416	18.4%	46.6%	34.9%	1,683
Males	82.6%	16.1%	1.3%	4,279	13.2%	45.8%	41.0%	1,698
African Americans	83.1%	16.0%	0.9%	5,373	22.7%	55.2%	22.1%	1,596
Whites	80.1%	18.2%	1.7%	2,022	9.2%	37.8%	53.0%	1,544
Algebra								
All students	52.3%	43.2%	4.5%	7,876	2.3%	43.8%	53.9%	3,497
Females	46.4%	48.5%	5.1%	3,416	2.4%	42.2%	55.3%	1,683
Males	56.8%	39.2%	4.0%	4,279	2.4%	45.1%	52.6%	1,698
African Americans	50.4%	44.5%	5.0%	5,373	2.6%	50.3%	47.2%	1,596
Whites	57.8%	39.3%	2.9%	2,022	2.3%	38.6%	59.1%	1,544
Measurement and Geometry								
All students	64.7%	34.7%	0.5%	7,876	6.5%	55.1%	38.4%	3,497
Females	63.3%	36.2%	0.6%	3,416	7.4%	56.7%	35.9%	1,683
Males	65.8%	33.7%	0.5%	4,279	5.6%	54.1%	40.3%	1,698
African Americans	66.5%	33.1%	0.4%	5,373	9.4%	69.3%	21.3%	1,596
Whites	60.7%	38.5%	0.8%	2,022	4.1%	43.5%	52.4%	1,544
Data Analysis and Probability								
All students	39.8%	57.6%	2.6%	7,876	3.6%	60.3%	36.1%	3,497
Females	37.5%	59.7%	2.7%	3,416	2.8%	61.9%	35.3%	1,683
Males	41.5%	55.9%	2.6%	4,279	4.3%	59.0%	36.7%	1,698
African Americans	39.8%	57.8%	2.4%	5,373	3.6%	69.6%	26.8%	1,596
Whites	38.5%	58.2%	3.3%	2,022	2.9%	50.8%	46.2%	1,544

* total number students in Level 1
** total number students in Levels 2, 3, and 4

TABLE 6.10

Spring 2005 HSAP Mathematics Operational Test: Content-Area Information (First Attempt)

Subgroup	Level 1				Level 2 and Above			
	Needs Improvement	*May need improvement*	*Adequate*	N1*	*Needs Improvement*	*May need improvement*	*Adequate*	N2**
Number and Operations								
All students	82.8%	16.3%	0.8%	12,590	6.8%	26.8%	66.4%	38,680
Females	83.6%	15.7%	0.6%	5,817	8.2%	29.4%	62.4%	19,753
Males	81.9%	17.0%	1.1%	6,591	5.2%	24.0%	70.8%	18,753
African Americans	84.3%	15.0%	0.7%	8,059	11.8%	38.2%	50.0%	12,167
Whites	79.1%	19.6%	1.3%	3,745	4.2%	21.0%	74.8%	24,509
Algebra								
All students	39.2%	57.1%	3.7%	12,590	0.7%	25.2%	74.1%	38,680
Females	32.2%	63.0%	4.8%	5,817	0.4%	23.7%	75.9%	19,753
Males	45.1%	52.2%	2.7%	6,591	0.9%	26.7%	72.4%	18,753
African Americans	38.4%	57.6%	4.0%	8,059	0.8%	32.6%	66.6%	12,167
Whites	41.0%	56.0%	3.0%	3,745	0.6%	21.4%	78.0%	24,509
Measurement and Geometry								
All students	53.5%	45.7%	0.8%	12,590	2.0%	32.7%	65.4%	38,680
Females	53.8%	45.6%	0.6%	5,817	2.2%	35.9%	61.9%	19,753
Males	53.0%	46.0%	1.0%	6,591	1.7%	29.2%	69.2%	18,753
African Americans	55.8%	43.5%	0.6%	8,059	3.4%	49.4%	47.2%	12,167
Whites	48.2%	50.6%	1.2%	3,745	1.2%	24.3%	74.5%	24,509
Data Analysis and Probability								
All students	32.6%	61.4%	6.0%	12,590	1.0%	25.8%	73.2%	38,680
Females	31.4%	62.5%	6.1%	5,817	1.2%	28.2%	70.7%	19,753
Males	33.3%	60.6%	6.0%	6,591	0.8%	23.2%	76.0%	18,753
African Americans	33.6%	61.1%	5.3%	8,059	1.8%	38.7%	59.5%	12,167
Whites	30.6%	61.7%	7.7%	3,745	0.6%	19.3%	80.1%	24,509

* total number students in Level 1
** total number students in Levels 2, 3, and 4

TABLE 6.11

Fall 2004 HSAP ELA Operational Test:

Content-Area Information (All Attempts)

Subgroup	Level 1				Level 2 and Above			
	Needs Improvement	*May need improvement*	*Adequate*	N1*	*Needs Improvement*	*May need improvement*	*Adequate*	N2**
Reading Process and Comprehension								
All students	59.7%	38.1%	2.2%	5,195	2.6%	44.9%	52.5%	3,917
Females	55.9%	42.2%	1.9%	1,727	2.8%	45.0%	52.2%	1,785
Males	61.6%	36.1%	2.3%	3,324	2.5%	45.2%	52.3%	2,011
African Americans	60.6%	37.6%	1.8%	3,478	3.9%	57.7%	38.3%	1,952
Whites	56.1%	40.7%	3.2%	1,247	1.3%	30.7%	67.9%	1,591
Analysis of Texts								
All students	63.9%	29.6%	6.5%	5,195	5.2%	31.1%	63.7%	3,917
Females	55.4%	35.8%	8.7%	1,727	3.4%	28.0%	68.6%	1,785
Males	68.1%	26.5%	5.4%	3,324	6.9%	34.0%	59.1%	2,011
African Americans	63.5%	30.4%	6.1%	3,478	5.9%	38.2%	55.9%	1,952
Whites	65.2%	28.1%	6.7%	1,247	4.7%	23.0%	72.3%	1,591
Word Study and Analysis								
All students	65.1%	32.3%	2.6%	5,195	10.3%	51.6%	38.0%	3,917
Females	61.7%	36.0%	2.3%	1,727	10.9%	51.3%	37.8%	1,785
Males	66.6%	30.6%	2.8%	3,324	10.0%	52.1%	37.9%	2,011
African American	65.5%	32.4%	2.1%	3,478	13.9%	61.3%	24.8%	1,952
White	62.3%	33.7%	4.0%	1,247	7.0%	40.9%	52.1%	1,591
Writing								
All students	78.0%	20.2%	1.8%	5,195	7.9%	37.4%	54.7%	3,917
Females	72.6%	25.2%	2.2%	1,727	5.2%	35.5%	59.4%	1,785
Males	80.9%	17.5%	1.6%	3,324	10.3%	39.8%	49.9%	2,011
African Americans	76.8%	21.2%	2.0%	3,478	9.5%	46.0%	44.5%	1,952
Whites	80.8%	17.5%	1.7%	1,247	6.0%	28.4%	65.6%	1,591
Research								
All students	52.8%	45.5%	1.6%	5,195	10.4%	67.2%	22.4%	3,917
Females	53.2%	45.8%	1.0%	1,727	12.0%	67.2%	20.8%	1,785
Males	52.6%	45.5%	1.9%	3,324	9.2%	67.0%	23.8%	2,011
African Americans	53.6%	44.9%	1.4%	3,478	13.9%	73.3%	12.8%	1,952
Whites	51.6%	46.3%	2.2%	1,247	6.5%	59.5%	34.0%	1,591

* total number students in Level 1
** total number students in Levels 2, 3, and 4

TABLE 6.12

Spring 2005 HSAP ELA Operational Test: Content-Area Information (First Attempt)

Subgroup	Level 1				Level 2 and Above			
	Needs Improvement	*May need improvement*	*Adequate*	N1*	*Needs Improvement*	*May need improvement*	*Adequate*	N2**
Reading Process and Comprehension								
All students	61.1%	36.4%	2.5%	7,440	1.0%	18.6%	80.4%	44,019
Females	56.1%	41.6%	2.3%	2,664	0.8%	16.6%	82.7%	22,965
Males	63.8%	33.7%	2.6%	4,668	1.3%	20.6%	78.0%	20,847
African Americans	59.4%	38.2%	2.3%	4,799	1.5%	27.7%	70.7%	15,549
Whites	63.5%	33.8%	2.7%	2,064	0.7%	13.0%	86.2%	26,287
Analysis of Texts								
All students	65.1%	32.2%	2.7%	7,440	3.8%	28.9%	67.3%	44,019
Females	67.9%	30.6%	1.6%	2,664	4.4%	30.4%	65.3%	22,965
Males	63.3%	33.4%	3.3%	4,668	3.2%	27.0%	69.8%	20,847
African Americans	66.7%	31.4%	1.9%	4,799	6.6%	42.5%	50.8%	15,549
Whites	61.5%	34.0%	4.6%	2,064	2.2%	20.6%	77.2%	26,287
Word Study and Analysis								
All students	51.5%	45.1%	3.4%	7,440	2.9%	39.6%	57.5%	44,019
Females	55.0%	43.0%	2.0%	2,664	3.6%	40.9%	55.4%	22,965
Males	49.2%	46.5%	4.3%	4,668	2.0%	38.0%	60.0%	20,847
African Americans	54.7%	43.1%	2.2%	4,799	5.6%	55.1%	39.4%	15,549
Whites	43.5%	50.3%	6.2%	2,064	1.3%	30.2%	68.6%	26,287
Writing								
All students	73.8%	23.7%	2.5%	7,440	2.5%	18.6%	78.9%	44,019
Females	66.6%	30.2%	3.2%	2,664	1.7%	16.3%	82.0%	22,965
Males	77.7%	20.1%	2.1%	4,668	3.4%	20.9%	75.7%	20,847
African Americans	74.1%	23.5%	2.4%	4,799	4.0%	28.2%	67.8%	15,549
Whites	71.8%	25.3%	3.0%	2,064	1.6%	12.6%	85.8%	26,287
Research								
All students	54.5%	43.1%	2.4%	7,440	5.7%	51.6%	42.7%	44,019
Females	51.5%	46.4%	2.1%	2,664	5.2%	50.8%	44.0%	22,965
Males	56.2%	41.3%	2.5%	4,668	6.3%	52.3%	41.3%	20,847
African Americans	52.8%	44.9%	2.3%	4,799	8.0%	62.4%	29.6%	15,549
Whites	59.2%	38.6%	2.2%	2,064	4.4%	45.1%	50.6%	26,287

* total number students in Level 1

** total number students in Levels 2, 3, and 4

Chapter 7

DESCRIPTIVE STATISTICS

Descriptive statistics of scale score distributions are presented in table 7.1. The scale score distributions are compared among all students, gender, and ethnic group categories in figures 1 through 4.

TABLE 7.1

HSAP Summary Statistics Overall and by Subgroups

Subgroup	Number	Scale Score	
		Mean	SD
Mathematics			
Fall 2004			
All students	11,373	192.7	19.9
Females	5,099	194.4	18.9
Males	5,977	191.0	20.4
African Americans	6,969	188.7	15.2
Whites	3,566	199.1	24.5
Spring 2005			
All students	51,270	220.2	28.6
Females	25,570	220.3	27.0
Males	25,344	220.3	30.1
African Americans	20,226	206.7	22.4
Whites	28,254	229.9	28.4
ELA			
Fall 2004			
All students	9,112	197.9	23.4
Females	3,512	202.8	23.0
Males	5,335	194.6	22.9
African Americans	5,430	193.0	19.1
Whites	2,838	206.9	27.2
Spring 2005			
All students	51,459	225.3	24.9
Females	25,629	228.2	23.7
Males	25,515	222.6	25.6
African Americans	20,348	213.9	21.6
Whites	28,351	233.9	23.5

FIGURE 1

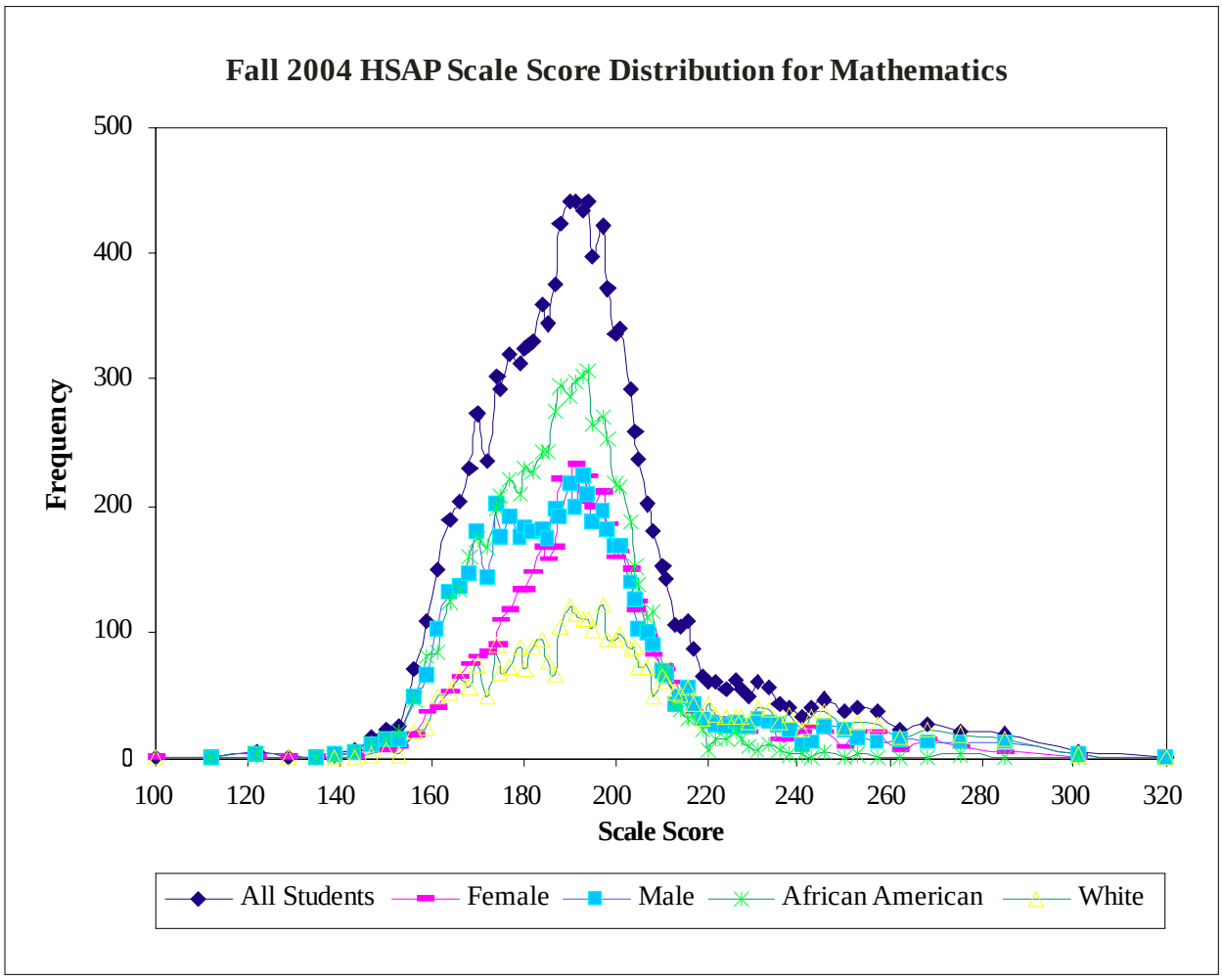

Fall 2004 HSAP Scale Score Distribution for Mathematics
Frequency
500
400
300
200
100
0
100
120
140
160
180
200
220
240
260
280
300
320
Scale Score
All Students
Female
Male
African American
White

FIGURE 2

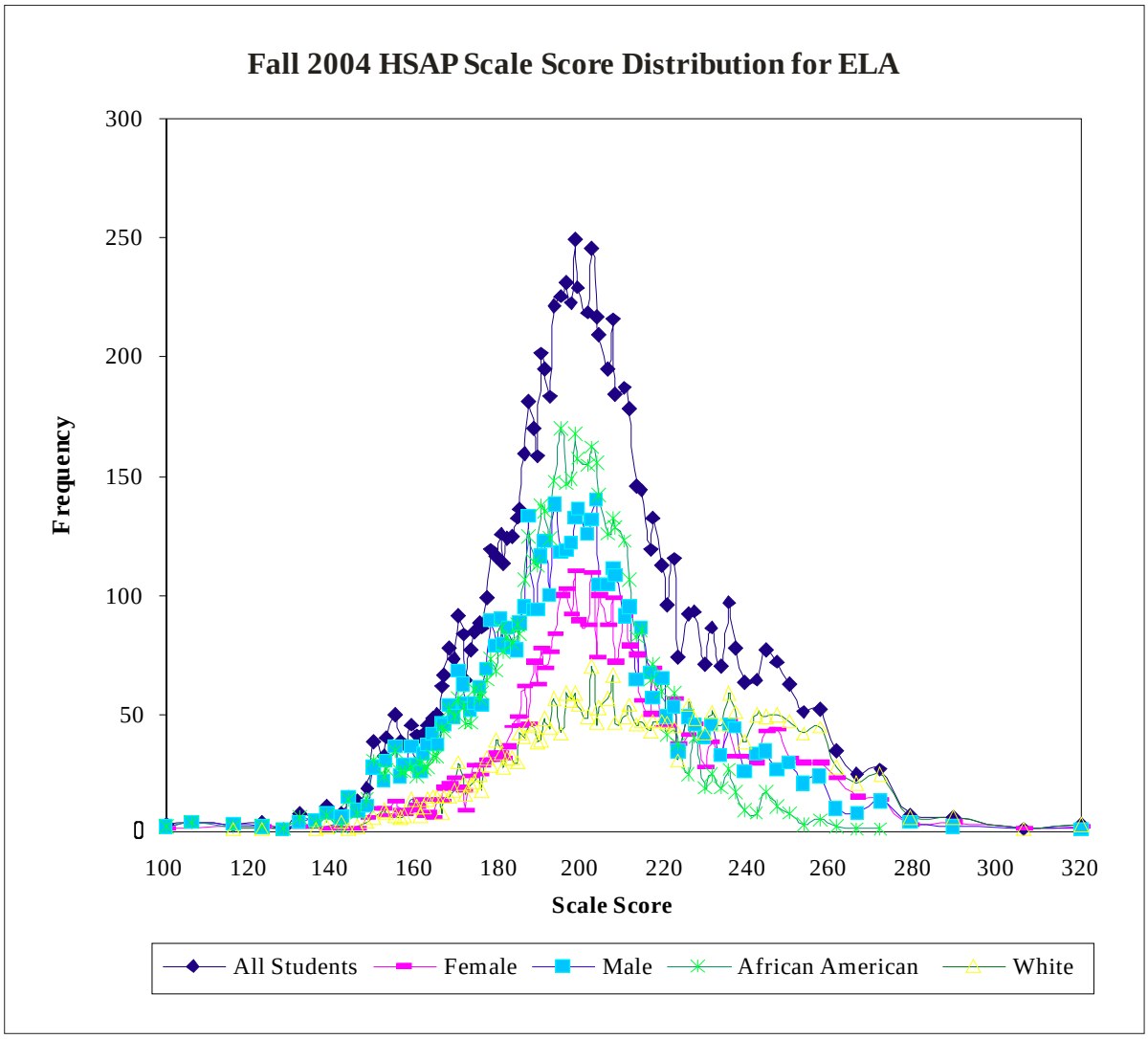

Fall 2004 HSAP Scale Score Distribution for ELA
Frequency
300
250
200
150
100
50
0
100
120
140
160
180
200
220
240
260
280
300
320
Scale Score
All Students
Female
Male
African American
White

FIGURE 3

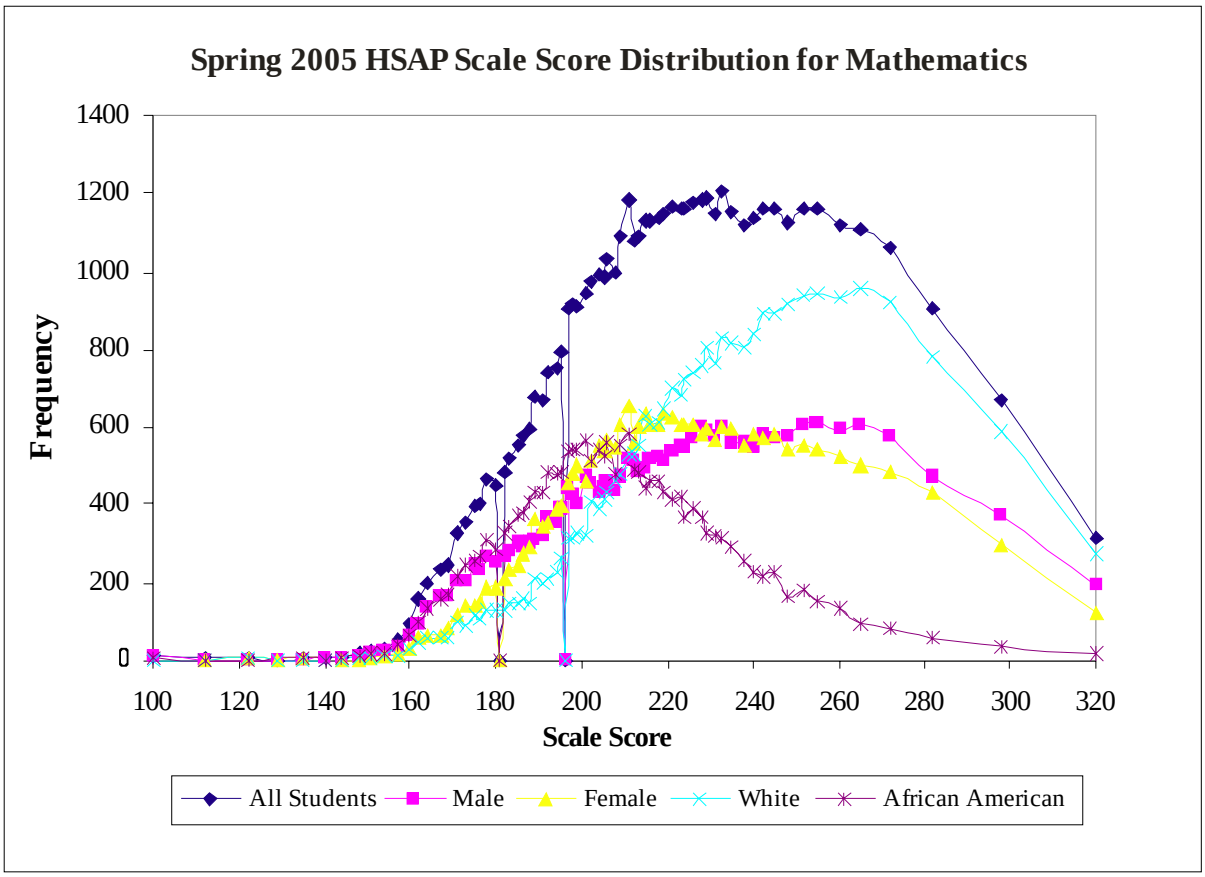
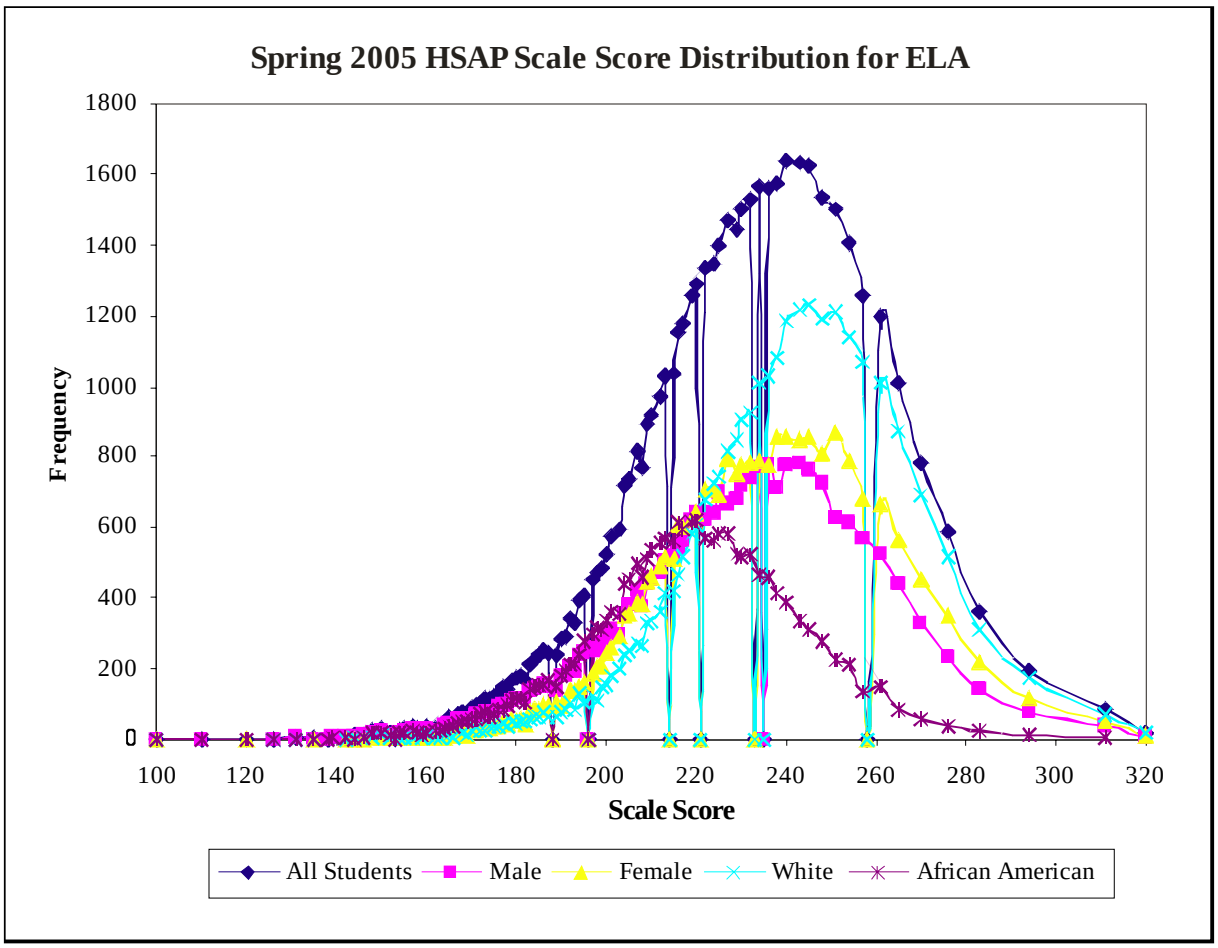

Spring 2005 HSAP Scale Score Distribution for Mathematics
Frequency
1400
1200
1000
800
600
400
200
0
100
120
140
160
180
200
220
240
260
280
300
320
Scale Score
All Students
Male
Female
White
African American

FIGURE 4


Spring 2005 HSAP Scale Score Distribution for ELA
Frequency
1800
1600
1400
1200
1000
800
600
400
200
0
100
120
140
160
180
200
220
240
260
280
300
320
Scale Score
All Students
Male
Female
White
African American

Chapter 8

RELIABILITY

In this chapter, three types of reliability indexes are presented: reliability of raw scores, overall standard error of measurement, conditional standard error of measurement, and decision consistency at each achievement level.

8.1 RELIABILITY OF RAW SCORES

For the HSAP assessments, the reliability coefficients were computed using stratified Cronbach's alpha. As mentioned, the HSAP assessments included mixed item types: multiple choice, constructed response, and extended response. Although there are various techniques for estimating the reliability of test scores with multiple item types or parts (Feldt and Brennan 1989; Lee and Frisbie 1999; Qualls 1995), studies indicate (Qualls 1995; Yoon and Young 2000) that the use of Cronbach's alpha underestimates the reliability of test scores for a test with mixed item types. The stratified coefficient alpha (Qualls 1995) is defined as

$$_{strat\,\alpha}\rho_{XX'} = 1 - \frac{\sum \sigma_{Y_j}^2 (1 - {}_{\alpha}\rho_{Y_jY_j'})}{\sigma_X^2}$$

where, σ_X^2 = the total score variance; $\sigma_{Y_j}^2$ = the score variance for a part-test j; ${}_{\alpha}\rho_{Y_jY_j'}$ = reliability of the part-test j.

Table 8.1 presents the reliability coefficients and standard errors of measurement for mathematics and ELA for all students and subgroups. The maximum possible score is 71 in mathematics and 96 in ELA.

TABLE 8.1

Reliability Coefficients and Standard Errors of Measurement for Raw Scores

	Fall 2004	Spring 2005
Mathematics		
Reliability	0.90	0.95
SEM	3.65	3.38
ELA		
Reliability	0.95	0.94
SEM	3.84	3.33

8.2 OVERALL AND CONDITIONAL STANDARD ERRORS OF MEASUREMENT

Table 8.2 presents the classical test-theory standard error of measurement (SEM) and the IRT-based conditional SEM at the scale score cutoff points. The 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. IRT-based conditional SEM at the scale score cutoff points are 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 cutoff points (Hambleton, Swaminathan, and Rogers 1991). Although classical SEM and IRT conditional SEM both serve the same role, the value of IRT-based conditional SEM varies with ability levels, whereas the classical SEM does not.

TABLE 8.2

Classical and Conditional Standard Errors of Measurement

Subject	Classical SEM	IRT-Based Conditional SEM		
		L2	L3	L4
Mathematics, fall 2004	6.2	5.6	6.1	7.7
Mathematics, spring 2005	6.6	5.6	6.0	7.7
ELA, fall 2004	5.2	5.5	6.3	7.7
ELA, spring 2005	5.6	5.4	6.2	7.5

Note: The SEM metric is in scale score points.

8.3 CONSISTENCY OF ACHIEVEMENT LEVELS

When student performance is reported in terms of achievement categories, a reliability index is computed in terms of the probabilities of consistent classification of students as specified in the standard 2.15 in *Standards for Educational and Psychological Testing* (AERA 1999). This index considers the consistency of classifications for the percentage of examinees that would, hypothetically, be classified in the same category on a second HSAP 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 8.3 presents a summary of agreements between the operational test classifications; that is, the percentages of students who are consistently classified in the same achievement levels on two equivalent administrations of the test.

TABLE 8.3

Consistency Indexes for Achievement Levels

Subject	Level 2	Level 3
Fall 2004		
Mathematics	90.0%	96.5%
ELA	92.2%	95.5%
Spring 2005		
Mathematics	94.3%	92.4%
ELA	95.2%	91.8%

Chapter 9

VALIDITY

Three types of validity evidence are reported in this section: test content, item fairness, and internal structure. Evidence on content validity is presented using the distribution of item content across content areas and the alignment of the fall 2004 and spring 2005 HSAP operational test items with reference to the state curriculum standards. Evidence on item fairness is examined with the information on differential item functioning (DIF). Evidence on internal structure is provided in correlations among content areas.

9.1 ITEM DISTRIBUTION ACROSS STRANDS

The HSAP operational test forms were constructed from the precalibrated item pools that were created based on the 2003 census field-test administration. These items measured the specific assessment standards that have been approved by the SDE. All items in the operational forms were reviewed by the Content Review Committee (CRC) and the Sensitivity Review Committee (SRC) and were approved by the SDE. The fall 2004 and spring 2005 HSAP test specifications are presented in section 4.2, above, in terms of distribution of score point values by content area.

9.2 ITEM DEVELOPMENT

All HSAP items were developed in alignment with the South Carolina curriculum standards and measurement guidelines. Various committees reviewed all items; only items reviewed by these committees and approved by the SDE were included in the operational forms. The embedded field-test items in ELA 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. After the SDE's review, the items were reviewed by the CRC and the SRC.

9.3 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 a pool of items that are fair to all students. All HSAP items were reviewed for bias and differential item functioning. The SRC reviewed the HSAP items for potential bias, including language that might disadvantage a group, might be considered offensive to members of a particular group, or might present obstacles to a group due to factors unrelated to content and processes specified in the standards.

After data were collected, the differential item functioning (DIF) statistics were produced for the statistical review. 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 *DIF* refers to statistical properties of an item in two equally able groups and is subject to later interpretation and judgment. Once an item is flagged for a significant DIF, judgment should be used to decide 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 procedures that the AIR selected for detecting DIF were the Mantel-Haenszel (MH) chi-square for dichotomous items (MC items) and Mantel's chi-square for polytomous items (CR and ER items). The AIR calculated the Mantel-Haenszel statistic (MH D-DIF) for MC items (Holland and Thayer 1988) and standardized mean difference (SMD) for CR items (Zwick, Donoghue, and Grima 1993) to measure the degree and magnitude of DIF.

The examinee group of interest is the *focal* group; the group to which performance on the item is being compared is the *reference* group. In this report, the focal groups for DIF were female and African American. Based on the DIF statistics, items were separated into one of three categories (Holland and Thayer 1988; Dorans and Holland 1993): negligible DIF (A), intermediate DIF (B), and large DIF (C). The items in category C, which exhibit significant DIF, are of primary concern.

For MC items, positive values of *delta* indicate that a given item is easier for the focal group, suggesting that the item favors the focal group. A negative value of *delta* indicates that a given item is more difficult for the focal group. Similarly, for CR items, a positive SMD value implies that, conditional on the matching variable (i.e., a total score), the focal group has a higher mean item score than the reference group, thereby favoring the focal group.

For MC items, 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 the delta value (Δ) is not significantly different from 0 or the absolute value of delta ($|\Delta|$) is less than or equal to 1.

For constructed-response items, the item classifications are based on the Mantel chi-square and the SMD index as follows:

- The item is classified as C category if the Mantel chi-square *p* value is less than .05 and the absolute value of SMD divided by standard deviation of the item score (i.e., |SMD/SD|) is larger than .25.
- The item is classified as B category if the Mantel chi-square *p* value is less than .05 and the absolute value of SMD divided by standard deviation of the item score (i.e., |SMD/SD|) is larger than .17.
- All other items will be classified as A category.

The number of items in DIF categories for the fall 2004 and spring 2005 mathematics and ELA operational items and ELA field-test items is summarized below in tables 9.1 through 9.3.

When items for the operational forms were selected, each item's statistics from the initial field test were reviewed and approved by the SDE. The 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 such items. For the fall 2004 operational forms, two multiple-choice items with C– (gender) and C+ (ethnicity) in mathematics and no items with a C category in ELA were included. For the spring 2005 operational forms, two multiple-choice items with C+ (gender and ethnicity) were included in mathematics; no items with a C category were included in ELA.

When the operational test data were analyzed, two multiple-choice items in ethnicity and one constructed-response item in gender exhibited C– and C+, respectively, in mathematics in fall 2004; one multiple-choice item exhibited C+ in gender in spring 2005. In ELA, two multiple-choice items exhibited C– in ethnicity and one constructed-response item exhibited C+ in gender in fall 2004; two multiple-choice items exhibited C– in gender and in ethnicity in spring 2005.

TABLE 9.1

Fall 2004 Summary of Differential Item Functioning for Mathematics and ELA Operational Items

Item Type	Reference Group	Focal Group	Total N of Items	DIF Classification		
				A	B	C
Mathematics						
Multiple choice	Male	Female	62	56	6	0
Multiple choice	White	Black	62	54	6	2
Constructed response	Male	Female	3	1	1	1
Constructed response	White	Black	3	3	0	0
ELA						
Multiple choice	Male	Female	60	57	3	0
Multiple choice	White	Black	60	54	4	2
Constructed response	Male	Female	2	1	0	1
Constructed response	White	Black	2	2	0	0
Extended response	Male	Female	8	8	0	0
Extended response	White	Black	8	7	1	0

TABLE 9.2

Spring 2005 Summary of Differential Item Functioning for Mathematics and ELA Operational Items

Item Type	Reference Group	Focal Group	Total N of Items	DIF Classification		
				A	B	C
Mathematics						
Multiple choice	Male	Female	62	54	7	1
Multiple choice	White	Black	62	56	6	0
Constructed response	Male	Female	3	3	0	0
Constructed response	White	Black	3	3	0	0
ELA						
Multiple choice	Male	Female	60	54	5	1
Multiple choice	White	Black	60	54	5	1
Constructed response	Male	Female	2	1	1	0
Constructed response	White	Black	2	2	0	0
Extended response	Male	Female	8	8	0	0
Extended response	White	Black	8	8	0	0

TABLE 9.3

Spring 2005 Summary of Differential Item Functioning for ELA Field-Test Items

Item Type	Reference Group	Focal Group	Total N of Items	DIF Classification		
				A	B	C
Multiple choice	Male	Female	80	64	13	3
Multiple choice	White	Black	80	77	2	1

9.4 CORRELATIONS AMONG REPORTING CATEGORIES

Reporting categories for mathematics include the following five areas: Algebra (AL), Number and Operations (NO), Measurement and Geometry (MG), Data Analysis and Probability (DP), and integrated responses (IR). ELA also includes five reporting categories: Reading Process and Comprehension (RC), Analysis of Texts (AT), Word Study and Analysis (WS), Research (RS), and Writing (WR). Tables 9.4 and 9.5 report the correlation matrices among the reporting category scores.

TABLE 9.4

Fall 2004 Correlations among Reporting Categories

Mathematics (N = 11,373)						ELA (N = 9,112)					
Reporting Category	**NO**	**AL**	**MG**	**DP**	**IR**	**Reporting Category**	**RC**	**AT**	**WS**	**WR**	**RS**
NO	1.00	0.67	0.64	0.55	0.64	**RC**	1.00	0.73	0.70	0.68	0.60
AL		1.00	0.62	0.54	0.63	**AT**		1.00	0.65	0.61	0.55
MG			1.00	0.57	0.67	**WS**			1.00	0.61	0.54
DP				1.00	0.55	**WR**				1.00	0.48
IR					1.00	**RS**					1.00

TABLE 9.5

Spring 2005 Correlations among Reporting Categories

Mathematics (N = 51,270)						ELA (N = 51,459)					
Reporting Category	**NO**	**AL**	**MG**	**DP**	**IR**	**Reporting Category**	**RC**	**AT**	**WS**	**WR**	**RS**
NO	1.00	0.76	0.76	0.72	0.76	**RC**	1.00	0.73	0.69	0.72	0.66
AL		1.00	0.75	0.69	0.73	**AT**		1.00	0.67	0.64	0.61
MG			1.00	0.72	0.76	**WS**			1.00	0.61	0.56
DP				1.00	0.70	**WR**				1.00	0.57
IR					1.00	**RS**					1.00

REFERENCES

AERA. 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 Robert L. Brennan. 1989. "Reliability." In *Educational Measurement*, edited by Robert L. Linn. 3rd ed. Washington, DC: American Council on Education.

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.

Lee, Guemin, and David A. Frisbie. 1999. "Estimating Reliability under a Generalizability Theory Model for Test Scores Composed of Testlets." *Applied Measurement in Education* 12, no. 3:237–55.

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.

Masters, Geofferey N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47:149–74.

Qualls, Audrey L. 1995. "Estimating the Reliability of a Test Containing Multiple Item Formats." *Applied Measurement in Education* 8, no. 2:111–20.

Rasch, Georg. 1980. *Probabilistic Models for Some Intelligence and Attainment Tests*. Rev. ed. Chicago: University of Chicago Press.

SDE. 2004a. *HSAP Test Administration Manual.* Columbia: South Carolina Department of Education.

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

———. 2005a. *HSAP Test Administration Manual.* Columbia: South Carolina Department of Education.

———. 2005b. *HSAP 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 Geoff Masters. 1982. *Rating Scale Analysis*. Chicago: MESA Press.

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

Yoon, Bokhee, and Michael J. Young. 2000. "Estimating the Reliability for Test Scores with Mixed Item Formats: Internal Consistency and Generalizability." Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Zwick, Rebecca, John R. Donoghue, and Angela Grima. 1993. "Assessment of Differential Item Functioning for Performance Tasks." *Journal of Educational Measurement* 30, no. 3:223–51.