

SCSL Digital Collections

South Carolina High School Assessment Program : English Language Arts and Mathematics 2008-09 operational test technical report

Item Type	Text
Publisher	South Carolina State Library
Rights	Copyright status undetermined. For more information contact, South Carolina State Library, 1500 Senate Street, Columbia, South Carolina 29201.
Download date	2024-10-06 03:27:43
Link to Item	http://hdl.handle.net/10827/30712

**South Carolina
High School Assessment Program**

**English Language Arts and Mathematics
2008–09 Operational Test Technical Report**



South Carolina
Department of Education

Together, we can.

Issued by the
South Carolina Department of Education

**Office of Assessment
Division of Accountability**

**Jim Rex
State Superintendent of Education**

CONTENTS

LIST OF TABLES.....	iv
CHAPTER 1. HSAP HISTORY AND OVERVIEW.....	1
1.1 Test Participation Requirements.....	1
1.2 Test Design and Structure.....	2
1.3 Technical Report Content.....	2
CHAPTER 2. STUDENT DEMOGRAPHICS.....	3
2.1 Student Participation.....	3
2.2 Accommodations and Modifications.....	7
Accommodations.....	7
Modifications.....	9
2.3 Student Questionnaires.....	10
CHAPTER 3. TEST ADMINISTRATION.....	11
3.1 Test Administration Window.....	11
3.2 Test Duration.....	11
3.3 Administration Manuals.....	12
3.4 Customized Materials.....	12
3.5 Pretest Workshops and Training.....	13
3.6 Materials Shipping and Return.....	13
3.7 Test Security.....	14
Secure Materials.....	15
CHAPTER 4. SCORING.....	16
4.1 Types of Items.....	16
Multiple Choice.....	16
Constructed Response.....	16
Extended Response.....	17
Graduation Express Automatic Rescore.....	17
4.2 Test Specifications.....	19
4.3 Scoring Process.....	20
4.4 Reader Reliability.....	20
4.5 Tested/Not Tested Flag.....	22
CHAPTER 5. TECHNICAL CHARACTERISTICS OF ITEMS.....	23
5.1 Item Nonresponse Rates.....	23
5.2 Classical Item Statistics.....	24
CHAPTER 6. ITEM CALIBRATION AND SCALING.....	25
6.1 Methodology and Software.....	25
6.2 Item Calibration.....	25
6.3 Scaling.....	25
6.4 Definition of Scoreability.....	26
6.5 Reporting of Zero and Perfect Score.....	26

6.6	Policy Definition of Achievement Levels.....	26
6.7	Cut Score for Achievement Levels.....	29
6.8	Content-Area Information.....	29
6.9	Percentage of Students in Each Achievement Level	31
6.10	Descriptive Statistics.....	44
CHAPTER 7. RELIABILITY		45
7.1	Reliability of Raw Scores	45
7.2	Overall and Conditional SEM.....	45
7.3	Consistency of Achievement Levels.....	46
CHAPTER 8. VALIDITY		47
8.1	Item Distribution across Strands.....	47
8.2	Item Development.....	47
8.3	Differential Item Functioning (DIF)	47
	Procedure	48
8.4	Correlations among Reporting Categories.....	50
REFERENCES		52

LIST OF TABLES

1.1	Students Scored through Graduation Express, Spring 2009	2
1.2	Number of Items.....	2
2.1	Fall 2008 Summary of Student Demographics in the HSAP Sample (All Attempts).....	4
2.2	Spring 2009 Summary of Student Demographics in the HSAP Sample (All Attempts)	5
2.3	Summer 2009 Summary of Student Demographics in the HSAP Sample (All Attempts)...	6
2.4	Percentages of Students with Accommodations in the Fall 2008 HSAP Administration (All Attempts).....	8
2.5	Percentages of Students with Accommodations in the Spring 2009 HSAP Administration (All Attempts)	8
2.6	Percentages of Students with Accommodations in the Summer 2009 HSAP Administration (All Attempts)	9
2.7	Percentages of Students with Modifications in the 2008–09 HSAP Administrations (All Attempts)	10
2.8	Number of Questions in the Student Questionnaire.....	10
3.1	2008–09 HSAP Test Administration Schedule	11
3.2	Percentage of Students by Test Duration: HSAP Fall 2008, Spring 2009, and Summer 2009 (All Attempts)	12
4.1	HSAP Extended-Response Scoring Algorithm for Papers with Scorable Responses	18
4.2	HSAP Extended-Response Scoring Algorithm for Papers with Condition Codes	18
4.3	HSAP Mathematics: Distribution of Score Point Values by Reporting Category	19
4.4	HSAP ELA: Distribution of Score Point Values by Reporting Category	19
4.5	Reader Reliabilities for Scoring HSAP Constructed-Response and Extended-Response Items.....	21
5.1	Percentages of Students Responding to Last and Second-to-Last HSAP Items	24
5.2	Summary of Classical Item Statistics for HSAP Mathematics and ELA.....	24
6.1	Description of Achievement Levels for the HSAP Mathematics Test.....	27
6.2	Description of Achievement Levels for the HSAP ELA Test	28
6.3	Rasch Ability and Scale Score Cut Scores for HSAP Achievement Levels	29
6.4	Cut Scores on the Rasch Ability Scale, Associated Standard Errors, and Confidence Intervals for HSAP Content-Area Classifications.....	30
6.5	Percentage of First Time Students Passing both Tests.....	31

6.6	Fall 2008 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	32
6.7	Spring 2009 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (First Attempt).....	33
6.8	Summer 2009 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	34
6.9	Fall 2008 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	35
6.10	Spring 2009 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (First Attempt).....	36
6.11	Summer 2009 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	37
6.12	Fall 2008 HSAP Mathematics Operational Test: Content-Area Information (All Attempts).....	38
6.13	Spring 2009 HSAP Mathematics Operational Test: Content-Area Information (First Attempt).....	39
6.14	Summer 2009 HSAP Mathematics Operational Test: Content-Area Information (All Attempts).....	40
6.15	Fall 2008 HSAP ELA Operational Test: Content-Area Information (All Attempts).....	41
6.16	Spring 2009 HSAP ELA Operational Test: Content-Area Information (First Attempt).....	42
6.17	Summer 2009 HSAP ELA Operational Test: Content-Area Information (All Attempts).....	43
6.18	HSAP Summary Statistics Overall and by Subgroups.....	44
7.1	Reliability Coefficients and SEM for HSAP Raw Scores.....	45
7.2	Classical and Conditional SEM for HSAP.....	46
7.3	Consistency Indexes for HSAP Achievement Levels	46
8.1	Fall 2008 Summary of DIF Classifications for HSAP Mathematics and ELA Items	49
8.2	Spring 2009 Summary of DIF Classifications for HSAP Mathematics and ELA Items ...	49
8.3	Summer 2009 Summary of DIF Classifications for HSAP Mathematics and ELA Items.	50
8.4	Fall 2008 HSAP Correlations among Reporting Categories.....	50
8.5	Spring 2009 HSAP Correlations among Reporting Categories	51
8.6	Summer 2009 HSAP Correlations among Reporting Categories	51

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 (NCLB) 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 NCLB 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. The first summer operational tests were administered in 2006.

The HSAP tests were developed by the South Carolina Department of Education (SCDE) and the American Institutes for Research (AIR) in 2002 and 2003. The test administration contractors have been AIR for spring 2003 through fall 2006, Pearson for spring 2007 through fall 2008, and Data Recognition Corporation (DRC) beginning in spring 2009.

1.1 TEST PARTICIPATION REQUIREMENTS

To be eligible to take the HSAP tests, students must be enrolled in a South Carolina public school, adult education program, or homeschool program approved by the local school board. Each of the three operational test administrations has distinct student-participation requirements:

- **Spring administration.** Students in the second year after their initial enrollment in the ninth grade take the HSAP in both mathematics and ELA. Students beyond the second year after their initial enrollment in the ninth grade take any HSAP test(s) they need in order to meet the requirement for a South Carolina high school diploma.
- **Fall administration.** Students beyond the second year after their initial enrollment in the ninth grade take any HSAP test(s) they need in order to meet the requirement for a South Carolina high school diploma.
- **Summer administration.** Following a remediation program in summer school, students who have not passed the exit examination and who are planning to graduate before the beginning of the next school year take any HSAP test(s) they need in order to meet the requirement for a South Carolina high school diploma.

Any student who fails either test will be scheduled to retake that test during the next scheduled administration for which he or she is eligible. A student who follows a normal progression of course work in high school has at least five opportunities—plus a sixth opportunity during the summer of his or her twelfth-grade year if necessary—to pass the exit examination. All accommodations and modifications available to students with disabilities and those with limited

English proficiency (LEP) for the spring administration are available to these students for the fall and summer administrations.

Beginning with the spring 2006 administration, students who were expected to graduate in the spring of the current year qualified for an expedited scoring process called Graduation Express. The number of students who qualified for Graduation Express is given in Table 1.1 below.

TABLE 1.1
Students Scored through Graduation Express, Spring 2009

Number of Students	
ELA	918
Math	1699

1.2 TEST DESIGN AND STRUCTURE

Table 1.2 contains the number of items on the forms administered. The item types are multiple choice (MC), constructed response (CR), and extended response (ER). The ELA form administered in Fall 2008 had only 58 multiple choice items since two items were treated as embedded field test items per SCDE request.

TABLE 1.2
Number of Items

	MC	CR	ER
ELA	60	2	1
Math	62	3	NA

1.3 TECHNICAL REPORT CONTENT

This technical report summarizes the results of statistical and psychometric analyses performed on the operational data for fall, spring, and summer of the current year's 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 and summer, the data summary in all chapters of this technical report includes all students who attempted the HSAP tests. For spring, the data in chapter 2, below, also include all students who attempted the HSAP tests; the data in other subsequent chapters include only those students who attempted the HSAP tests for the first time.

Chapter 2

STUDENT DEMOGRAPHICS

2.1 STUDENT PARTICIPATION

For all HSAP administrations, demographic data were collected on each student. These data included the categories of gender, race/ethnicity, grade, English language proficiency, lunch program eligibility, disability status, and migrant status. All data are based on students in the regular schools only; students in adult education and district-approved home schools were excluded. For clarity, adult education and homeschooled students were not included in statewide aggregate reports.

On the following pages, tables 2.1 through 2.3 report the demographic distributions. The “Invalid” category in these tables includes blanks and multiple marks. The fall pre-ID file contained data on students who did not pass the HSAP the previous spring. Because most students change grade level from spring to fall, all fall values for the variable “Grade” were taken from the hand-gridded information. The high *invalid* rate for the “Grade” category is due to the fact that some students and test administrators did not grid the grade field.

Table 2.1
Fall 2008 Summary of Student Demographics
in the HSAP Sample (All Attempts)

Demographic Category	Mathematics		ELA	
	N	%	N	%
All Students	11248	100.0	8886	100.0
Gender				
Female	5090	45.3	3523	39.6
Male	5734	51.0	5171	58.2
Invalid	424	3.8	192	2.2
Ethnicity				
African American	6419	57.1	4660	52.4
Asian/Pacific Islander	610	5.4	650	7.3
Hispanic	28	0.2	22	0.2
American Indian	246	2.2	204	2.3
White	118	1.0	154	1.7
Other	3423	30.4	2988	33.6
Invalid	404	3.6	208	2.3
Grade				
9	617	5.5	609	6.9
10	1349	12.0	1184	13.3
11	3006	26.7	2685	30.2
12	1372	12.2	1109	12.5
Invalid	4904	43.6	3299	37.1
ESL*				
No	3612	32.1	3349	37.7
Yes	7370	65.5	5221	58.8
Unknown	266	2.4	316	3.6
Lunch Program				
No free/reduced lunch	2254	20.0	2038	22.9
Free lunch	2717	24.2	2591	29.2
Reduced lunch	259	2.3	222	2.5
Unknown	6018	53.5	4035	45.4
IEP**				
No	9397	83.5	7375	83.0
Yes	1851	16.5	1511	17.0
Unknown				
Migrant				
No	5049	44.9	4648	52.3
Yes	6196	55.1	4229	47.6
Unknown	3	0.0	9	0.1
Attempt				
1st	3248	28.9	3242	36.5
2nd	5654	50.3	4114	46.3
3rd	1094	9.7	638	7.2
4th or more	977	8.7	622	7.0

* English as a second language

** individualized education program

Table 2.2
Spring 2009 Summary of Student Demographics
in the HSAP Sample (All Attempts)

Demographics	Mathematics		ELA	
	N	%	N	%
All Students	58592	100.0	56428	100.0
Gender				
Female	29234	49.9	27877	49.4
Male	29216	49.9	28453	50.4
Invalid	142	0.2	98	0.2
Ethnicity				
African American	24531	41.9	22837	40.5
Asian/Pacific Islander	649	1.1	669	1.2
Hispanic	2448	4.2	2474	4.4
American Indian	125	0.2	121	0.2
White	29838	50.9	29397	52.1
Other	813	1.4	789	1.4
Invalid	188	0.3	141	0.2
Grade				
9	6535	11.2	6442	11.4
10	47396	80.9	46924	83.2
11	3223	5.5	2185	3.9
12	1342	2.3	804	1.4
Invalid	96	0.2	73	0.1
ESL*				
Parent Waiver	96	0.2	100	0.2
Pre-Functional - Advanced	1378	2.4	1470	2.6
Initially English Proficient	71	0.1	72	0.1
Title III Exited	208	0.4	207	0.4
English Speaker I	202	0.3	200	0.4
English Speaker II	55436	94.6	53321	94.5
All Others	1201	2.0	1058	1.9
Lunch Program				
No free/reduced lunch	30310	51.7	29426	52.1
Free lunch	24155	41.2	23031	40.8
Reduced lunch	4127	7.0	3971	7.0
IEP**				
No	50073	85.5	48599	86.1
Yes	8519	14.5	7829	13.9
Migrant				
No	58578	100.0	56411	100.0
Yes	14	0.0	17	0.0
Attempt				
1st	53323	91.0	53088	94.1
2nd	2035	3.5	1408	2.5
3rd	2246	3.8	1356	2.4
4th or more	987	1.7	575	1.0

* English as a second language

** individualized education program

Table 2.3
Summer 2009 Summary of Student Demographics
in the HSAP Sample (All Attempts)

Demographics	Mathematics		ELA	
	N	%	N	%
All Students	203	100.0	82	100.0
Gender				
Female	109	53.7	40	48.8
Male	76	37.4	38	46.3
Invalid	18	8.9	4	4.9
Ethnicity				
African American	152	74.9	58	70.7
Asian/Pacific Islander	0	0.0	3	3.7
Hispanic	2	1.0	5	6.1
American Indian	1	0.5	0	0.0
White	22	10.8	9	11.0
Other	3	1.5	1	1.2
Invalid	23	11.3	6	7.3
Grade				
10	3	1.5	1	1.2
11	6	3.0	4	4.9
12	177	87.2	72	87.8
Invalid	17	8.4	5	6.1
ESL*				
Pre-Functional - Advanced	2	1.0	6	7.3
English Speaker II	73	36.0	33	40.2
All others	128	63.1	43	52.4
Lunch Program				
No free/reduced lunch	130	64.0	45	54.9
Free lunch	67	33.0	34	41.5
Reduced lunch	6	3.0	3	3.7
IEP**				
No	157	77.3	60	73.2
Yes	46	22.7	22	26.8
Migrant				
No	203	100.0	82	100.0
Yes	0	0.0	0	0.0
Attempt				
1st	11	5.4	6	7.3
2nd	2	1.0	0	0.0
3rd	11	5.4	5	6.1
4th or more	179	88.2	71	86.6

* English as a second language

** individualized education program

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)* (SCDE 2008a and 2009a). 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 accommodations include 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, large-print, and braille test booklets for both tests.

On the following pages, tables 2.4 through 2.6 present summaries of accommodations by the percentages of those students who were administered the test with one or more accommodations. (The column totals exceed 100 because some students received accommodations in more than one category.)

TABLE 2.4
Percentages of Students with Accommodations in the
Fall 2008 HSAP Administration (All Attempts)

Accommodation	Mathematics		ELA	
	Regular Form (N=11,222)	Customized Form (N=26)	Regular Form (N=8,861)	Customized Form (N=25)
Setting	0.7	73.1	0.6	84.0
Presentation	--	--	--	92.0
Timing	1.1	76.9	1.2	84.0
Schedule	15.5	96.2	16.7	96.0
Response options	1.2	69.2	1.0	80.0
Loose-leaf	6.5	--	1.8	--
Large-print	5.6	--	0.8	--
Spelling dictionary	0.7	88.5	0.4	80.0
Audiocassette	0.2	--	--	--
Oral script	0.1	53.8	0.2	--
Signed administration	0.7	88.5	0.4	80.0
Braille	--	--	--	--
Other	0.2	53.8	0.2	--

TABLE 2.5
Percentages of Students with Accommodations in the
Spring 2009 HSAP Administration (All Attempts)

Accommodation	Mathematics		ELA	
	Regular Form (N=58,538)	Customized Form (N=54)	Regular Form (N=56,373)	Customized Form (N=55)
Setting	8.1	92.6	8.3	96.4
Presentation	5.7	92.6	--	--
Timing (IEP)	0.4	64.8	0.5	65.5
Schedule (IEP)	0.1	61.1	0.1	74.5
Response options	0.2	11.1	0.5	61.8
Loose-leaf	0.0	--	0.0	--
Large-print	0.1	--	0.1	--
Spelling dictionary	--	--	0.5	54.5
Supplemental Materials	0.1	63.0	0.1	67.3
Audiocassette	3.4	3.7	--	--
Oral script	3.0	1.9	--	--
Signed administration	--	83.3	--	--
Braille	--	13.0	--	9.1
Bilingual Dictionary	0.9	--	1.2	--
Directions Translation	0.1	--	0.1	--
Individual/Small Group	0.9	5.6	0.9	5.5
Timing (ESL)	0.0	3.7	0.1	3.6
Schedule (ESL)	0.0	3.7	0.0	5.5
Other	0.0	--	0.0	--

TABLE 2.6
Percentages of Students with Accommodations in the
Summer 2009 Accommodations (All Attempts)

Accommodation	Mathematics		ELA	
	Regular Form (N=203)	Customized Form (N=0)	Regular Form (N=82)	Customized Form (N=0)
Setting	12.3	--	19.5	--
Presentation	18.2	--	--	--
Response options	0.5	--	1.2	--
Spelling dictionary	--	--	1.2	--
Audiocassette	12.3	--	--	--
Oral script	12.3	--	--	--
Bilingual Dictionary	1.0	--	7.3	--
Individual/Small Group	--	--	3.7	--
Timing (ESL)	--	--	1.2	--
Schedule (ESL)	--	--	1.2	--

Modifications

The term *modification* refers to a change in the testing environment, procedures, or presentation that compromises 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 ELA test modifications included oral administration, signed administration, alternative scoring for extended-response 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 2.7, below, presents summaries of modifications by percentages (again, the column totals may exceed 100 percent because some students received modifications in more than one category).

TABLE 2.7
Percentages of Students with Modifications
in the 2008–09 HSAP Administrations (All Attempts)

Modification	Regular Form	Customized Form
<hr/>		
Fall 2008	(N=8,861)	(N=25)
<hr/>		
Alternative scoring	4.3	92.0
Extended writing option	1.0	80.0
Audiocassette	7.7	--
Oral administration	6.3	--
Signed administration	--	100.0
<hr/>		
	Regular Form	Customized Form
Spring 2009	(N=56,373)	(N=55)
<hr/>		
Alternative scoring	2.0	81.8
Extended writing option	0.6	61.8
Audiocassette	3.7	--
Oral administration	5.9	1.8
Signed administration	--	90.9
<hr/>		
	Regular Form	Customized Form
Summer 2009	(N=82)	(N=0)
<hr/>		
Alternative scoring	6.1	--
Audiocassette	18.3	--
Oral administration	26.8	--
<hr/>		

2.3 STUDENT QUESTIONNAIRES

In fall 2008 and prior HSAP administrations, students answered questions regarding test difficulty, classroom activities, and (for mathematics only) calculator use. Student questionnaires were discontinued beginning in Spring 2009. Table 2.8 contains the number of questions on the fall 2008 student questionnaire.

TABLE 2.8
Number of Questions in the Student Questionnaire

	ELA	Math	Customized Math
Fall 2008	11	12	12

Chapter 3

TEST ADMINISTRATION

3.1 TEST ADMINISTRATION WINDOW

The HSAP ELA operational tests for fall and spring were conducted in two sessions over two days. The mathematics tests were conducted in one day. For the summer administration, school districts were responsible for identifying the test dates for each subject within the three-day window. There were no makeup testing days for the summer administration.

TABLE 3.1
2008–09 HSAP Test Administration Schedule

Fall 2008		Spring and Summer 2009	
Date	Test	Date	Test
Oct. 21	ELA (day 1)	Apr. 21	ELA (day 1)
Oct. 22	ELA (day 2)	Apr. 22	ELA (day 2)
Oct. 23	Mathematics	Apr. 23	Mathematics
Oct. 24 - Oct.31	Makeup tests window	Apr. 24-May 1	Makeup tests window
		Jul. 21-23	Summer tests window

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 TEST DURATION

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.

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. Table 3.2 reports the breakdowns by percentages. “Invalid” refers to blank or multiple responses. Total testing times for students whose responses fell into this category could not be calculated.

Table 3.2
Percentage of Students by Test Duration:
HSAP Fall 2008, Spring 2009, and Summer 2009 (All Attempts)

Time Taken	Fall 2008			Spring 2009			Summer 2009		
	Math % (N=11,248)	ELA % (N=8,886)		Math % (N=58,592)	ELA % (N=56,428)		Math % (N=203)	ELA % (N=82)	
		Session 1	Session 2		Session 1	Session 2		Session 1	Session 2
15 min	0.5	1.8	0.6	0.2	0.5	0.3	0.5	0.0	0.0
30 min	1.5	10.0	4.0	0.5	3.6	1.4	0.0	2.4	1.2
45 min	6.0	19.2	11.2	2.3	11.7	8.6	2.0	2.4	2.4
1 hr	14.4	21.1	19.0	8.9	20.4	22.5	0.0	9.8	7.3
1 hr 15 min	16.2	14.6	15.9	16.3	20.0	23.3	4.4	12.2	14.6
1 hr 30	15.1	9.4	13.0	19.0	15.8	16.2	6.9	15.9	14.6
1 hr 45	11.3	5.8	8.9	15.8	10.2	9.9	10.3	12.2	14.6
2 hr	9.7	4.5	7.7	12.4	6.4	6.1	12.8	15.9	9.8
2 hr 15 min	6.4	2.9	4.3	7.9	3.5	3.3	8.9	4.9	9.8
2 hr 30 min	4.3	1.4	3.0	5.0	1.9	1.9	11.3	8.5	7.3
2 hr 45 min	2.9	0.9	2.1	2.8	1.1	0.9	8.9	1.2	7.3
3 hr +	5.9	2.5	4.1	4.8	2.0	1.8	27.6	14.6	6.1
Invalid	5.7	5.9	6.3	4.1	3.0	3.9	6.4	0.0	4.9

3.3 ADMINISTRATION MANUALS

Testing contractors (Pearson in fall 2008 and DRC in spring and summer 2009) worked with SCDE staff to produce administration manuals for the test. Two types of manuals were produced: the *HSAP Test Administration Manual (TAM)* and the *HSAP District Test Coordinator's Supplement (SCDE 2008b and 2009b)*. The supplement included 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. In addition, the summer 2009 supplement (SCDE 2009c) was produced to provide specific information for the summer 2009 administration and was used in conjunction with the information provided in the spring 2009 manuals.

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.

- Form A 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.
- Form A 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.

- Form C 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.
- Form A oral administration scripts and audio CDs 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. Audio CDs were used for students testing individually or in small-group settings.
- Form C Braille scripts and audiotapes/CDs (tapes for fall 2008 and spring 2009, CDs for summer 2009) were produced for testing visually impaired students.
- Form C sign language videotapes were also produced and included the signed test directions, test questions, and some answer choices. The videotapes were produced in two languages: American Sign Language (ASL) and Pidgin Signed English (PSE). Signed Exact English (SEE) video tapes were used for the last time in fall 2008.

3.5 PRETEST WORKSHOPS AND TRAINING

Web-based pretest workshops were held to train DTCs and other district staff. The DTCs were invited and could bring additional representatives to the workshop. SCDE and Pearson staff trained the district staff in fall 2008. SCDE and DRC staff trained the district staff in spring and summer 2009.

Contractor staff discussed the HSAP manuals, reviewed test security procedures, and provided other pertinent information, including an overview of the instructions for administering tests to students with disabilities. Special focus was given to new procedures as well as any recent changes in procedure.

In fall 2008, the *DTC Supplements* and *TAMs* were mailed to the DTCs two weeks before the workshops and were also handed out to the DTCs during the workshop. The PowerPoint presentations were posted to the SCDE Web site and e-mailed to DTCs.

In spring 2009, the *TAMs* were mailed to the DTCs two weeks before the workshops, as well as posted online. The *DTC Supplements* were posted online the same day the *TAMs* were mailed to the districts. In addition, the PowerPoint presentations were posted online. In summer 2009, *HSAP Summer Supplements* were posted online; spring *TAMs* were reused by the districts.

3.6 MATERIALS SHIPPING AND RETURN

Test materials were shipped to the district offices by DRC to arrive at least two weeks before testing. Each school's shipment was boxed individually and labeled with the number of boxes shipped for that school. The DRC 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 summer 2009 shipment included a 5 percent overage for the districts and a 3 percent overage for the schools.

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.

The *DTC Supplement* 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 DRC. For spring testing, an additional shipment was made for Graduation Express students.

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 STCs. 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 instructed to walk around the room during testing to check that students were marking their answers in the appropriate areas of the test booklets/answer documents. It was permissible to alert students if they were marking their answers in the wrong areas. However, it was not permissible to read test items or students' responses.

Following the test administration and the return of materials, DRC 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 DRC 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 the *TAM*. DTCs were instructed to require all persons with access to test materials to sign test security agreements if not on file for the current school year. This necessity was stressed repeatedly in the manuals and during the pretest workshops.

Chapter 4

SCORING

The criteria used to score HSAP 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 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 on 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 answers to constructed-response items 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 readers. To produce a final score, the two scores were processed according to the scoring algorithms shown in tables 4.1 and 4.2, on the following page.

Graduation Express Automatic Rescore

The regular appeal process does not allow hand-scored responses of Graduation Express students to be rescored in time for graduation. Consequently, for the spring administrations of 2006, 2007, and 2008, all Graduation Express students’ CR and ER responses were automatically rescored. The higher score was used to calculate the student’s final score.

In spring 2009, the automatic rescore procedure for Graduation Express students was changed to apply only to graduating seniors who initially scored at level 1. The rescored response value was then used to calculate the student’s final score.

TABLE 4.1
HSAP Extended-Response Scoring Algorithm for Papers with Scorable Responses

Rule	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	NA	F = R1 + R2
2	R1 = 1–4	R2 = 1–4 and is adjacent to R1	none	NA	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	NA	BR = R1 = R2	NA	F = BR + R1
8	R1 = 1–4	R2 = R1	NA	BR is adjacent to R1 and R2	NA	F = BR + R1
9	R1 = 1–4	R2 = R1	NA	BR is nonadjacent to R1 and R2	NA	F = BR + BR
10	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1	NA	BR = R1 and adjacent to R2	NA	F = BR + R1
11	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1	NA	BR = R2 and adjacent to R1	NA	F = BR + R2
12	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1	NA	BR is adjacent to R1 and discrepant to R2	NA	F = BR + R1
13	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1	NA	BR is adjacent to R2 and discrepant to R1	NA	F = BR + R2
14	R1 = 1–4	R2 = 1–4 and R2 is adjacent to R1	NA	BR is nonadjacent to R1 and R2	NA	F = BR + BR

TABLE 4.2
HSAP Extended-Response 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	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 HSAP test specifications for mathematics and ELA are shown in tables 4.3 and 4.4, below. As noted previously, the HSAP assessments include 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
HSAP Mathematics: Distribution of Score Point Values by Reporting Category

Fall 2008, Spring and Summer 2009	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	16	—
Constructed-response points	—	—	—	—	9

TABLE 4.4
HSAP ELA: Distribution of Score Point Values by Reporting Category

	Reading Process and Comprehension	Analysis of Texts	Word Study and Analysis	Research	Writing
Fall 2008					
Percentage	26%	18%	9%	7%	40%
Multiple-choice points	18	17	8	7	8
Constructed-response points	6	--	--	--	--
Extended-response points	--	--	--	--	30
Spring 2009					
Percentage	28%	17%	8%	7%	40%
Multiple-choice points	21	16	8	7	8
Constructed-response points	6	--	--	--	--
Extended-response points	--	--	--	--	30
Summer 2009					
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

Pearson conducted scoring for all items in fall 2008. The multiple-choice items were scored by Pearson's electronic scanning system; constructed-response (CR) and extended-response (ER) items were scored by trained personnel at Pearson's Mesa, AZ scoring site. SCDE staff monitored the initial scoring. Throughout the scoring process, the contractor posted the performance of each reader (reader-reliability statistics) online daily.

DRC scored all items for spring and summer 2009. The multiple-choice items were scored by DRC's electronic scanning system; CR and ER items were scored by trained personnel at DRC's Austin, TX scoring site. SCDE staff were on-site during the first week of training in spring 2009. They oversaw both the training of reader leaders and the initial training of all readers. Throughout the scoring process, the contractor posted the performance of each reader (reader-reliability statistics). SCDE staff did not oversee training or scoring in summer 2009.

Before starting to score the live CR and ER items, readers have 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 85 percent adjacent on 2 of 3 sets with 20 papers in each set
- ELA CR: 90 percent exact on 2 of 3 sets with 20 papers in each set
- Math CR: 90 percent exact on 2 of 3 sets with 20 papers in each set.

Before qualifying and throughout the 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 SCDE and contractor content specialists and scoring directors. This quality check was "blind" in that readers did not know they were scoring a validity paper. In addition to validity, quality was monitored through the use of reader reliability and score point distribution reports. Reader agreement was checked on a regular basis: every twenty papers for the extended-response item and every sixty papers for CR items.

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 table 4.5, on the following page.

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
Reader Reliabilities for Scoring HSAP
Constructed-Response and Extended-Response Items

Items	N	Percentage of Perfect Agreement	Percentage of Perfect and Adjacent Agreement
Mathematics			
Fall 2008 (All Attempts)			
CR1	1,695	89.9	98.2
CR2	1,691	89.6	98.8
CR3	1,688	87.7	99.4
Spring 2009 (First Attempts)			
CR1	8,878	96.6	100.0
CR2	8,971	96.3	100.0
CR3	8,879	93.7	99.9
Summer 2009 (All Attempts)			
CR1	39	92.3	100.0
CR2	32	96.9	100.0
CR3	37	91.9	100.0
ELA			
Fall 2008 (All Attempts)			
CR1	1,334	74.1	99.0
CR2	1,356	77.2	98.5
ER content and development	8,886	65.2	98.6
ER organization	8,886	63.9	97.9
ER voice	8,886	64.8	98.5
ER convention	8,886	55.0	92.5
Spring 2009 (First Attempts)			
CR1	8,807	83.1	99.8
CR2	9,107	85.2	99.7
ER content and development	53,088	74.2	99.0
ER organization	53,088	73.0	98.8
ER voice	53,088	77.5	99.8
ER convention	53,088	75.0	98.4
Summer 2009 (All Attempts)			
CR1	15	100.0	100.0
CR2	14	85.7	100.0
ER content and development	82	85.4	100.0
ER organization	82	87.8	100.0
ER voice	82	87.8	100.0
ER convention	82	78.0	100.0

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 item.

Chapter 5

TECHNICAL CHARACTERISTICS OF ITEMS

This chapter reports the results of item analyses based on classical test theory (CTT). 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. It was decided to exclude from the CTT item analyses those students who had used customized materials and those who had received the alternative scoring rubric modification. These students were also excluded during the item calibration conducted prior to building pre-equated forms. This calibration was conducted by the AIR in coordination with the SCDE.

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, on the following page, 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 (question 65) 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 (question 64) for mathematics, item 13 in fall, item 16 in spring, and item 13 in summer in session 1 and item 59 in fall, spring, and summer 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. (Question numbers above are for Form A only.)

TABLE 5.1
Percentages of Students Responding to Last and Second-to-Last HSAP Items

Subject	Last Item	Second-to-Last Item
Fall 2008 (All Attempts)		
Mathematics	91.9 (CR)	91.6 (CR)
ELA Session 1	98.0 (MC)	98.6 (MC)
ELA Session 2	98.5 (MC)	98.5 (MC)
Spring 2009 (First Attempts)		
Mathematics	93.2 (CR)	96.5 (CR)
ELA Session 1	99.5 (MC)	99.5 (MC)
ELA Session 2	99.6 (MC)	99.6 (MC)
Summer 2009 (All Attempts)		
Mathematics	96.6 (CR)	97.5 (CR)
ELA Session 1	100.0 (MC)	100.0 (MC)
ELA Session 2	100.0 (MC)	100.0 (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 items. For constructed-response and extended-response items, the p -value was computed as the ratio of the item mean to the item's maximum possible score. 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 score was excluded from the total raw score.

TABLE 5.2
Summary of Classical Item Statistics for HSAP Mathematics and ELA

Item Type/Content Area	Number of Items	p -value	Point-Biserial Correlation	Number of Items	p -value	Point-Biserial Correlation	Number of Items	p -value	Point-Biserial Correlation
	Fall 2008 (All Attempts)			Spring 2009 (First Attempts)			Summer 2009 (All Attempts)		
Mathematics									
Multiple-choice	62	0.51	0.36	62	0.69	0.43	62	0.49	0.19
Constructed-response	3	0.36	0.61	3	0.56	0.70	3	0.30	0.42
Number and Operations	16	0.61	0.75	16	0.75	0.43	16	0.60	0.24
Algebra	19	0.54	0.72	19	0.70	0.42	19	0.49	0.19
Measurement and Geometry	19	0.40	0.75	19	0.62	0.46	19	0.37	0.14
Data Analysis and Probability	8	0.52	0.70	8	0.73	0.43	8	0.60	0.17
ELA									
Multiple-choice	58	0.57	0.37	60	0.73	0.36	60	0.53	0.22
Constructed-response	2	0.34	0.57	2	0.54	0.55	2	0.38	0.43
Extended-response	1	0.73	0.65	1	0.81	0.74	1	0.74	0.57
Reading Process and Comprehension	20	0.54	0.80	23	0.74	0.37	21	0.55	0.24
Analysis of Texts	17	0.51	0.73	16	0.75	0.35	17	0.57	0.27
Word Study and Analysis	8	0.42	0.65	8	0.72	0.36	8	0.49	0.23
Research	7	0.65	0.64	7	0.63	0.36	8	0.40	0.14
Writing	9	0.55	0.72	9	0.72	0.41	9	0.57	0.24

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 both mathematics and ELA, the equated HSAP operational test forms were constructed from the precalibrated item pool; therefore, the raw-score-to-scale-score conversion tables for the operational forms were created before the tests were administered.

6.3 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.

The following process is used to convert Rasch ability scores to scale scores:

Step 1: A linear transformation is applied to the Rasch scores ($\hat{\theta}$), such that the Level 2 cut score (SS_c) equals 200 and the standard deviation of scales scores (B) is 25,

$$SS = SS_c + B \left[\frac{\hat{\theta} - \theta_c}{\sigma_{\hat{\theta}}} \right]$$
, where the Rasch passing 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: Noninteger scale scores are rounded down to whole numbers.

Step 3: Scale scores less than 100 and greater than 320 are reported as 100 and 320, respectively.

6.4 DEFINITION OF SCOREABILITY

A student is considered “tested” if he or she has answered at least one question. All tested students’ item responses are scored. All omits and not-reached items are counted as incorrect and scored as a zero.

6.5 REPORTING OF ZERO AND PERFECT SCORE

In item response theory (IRT) maximum-likelihood ability estimation methods, zero and perfect scores are assigned the value of negative and positive infinity, respectively. 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.

6.6 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 on the following pages in tables 6.1 and 6.2. The AIR outlined the details of the standard-setting process in its 2004 report to the SCDE, “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	<p>The Level 4 student</p> <ul style="list-style-type: none"> • 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	<p>The Level 3 student</p> <ul style="list-style-type: none"> • 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	<p>The Level 2 student</p> <ul style="list-style-type: none"> • 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	<p>The Level 1 student</p> <ul style="list-style-type: none"> • 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	<p>The Level 4 student</p> <ul style="list-style-type: none"> • 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	<p>The Level 3 student</p> <ul style="list-style-type: none"> • 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	<p>The Level 2 student</p> <ul style="list-style-type: none"> • 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	<p>The Level 1 student</p> <ul style="list-style-type: none"> • 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

6.7 CUT SCORES FOR ACHIEVEMENT LEVELS

The cut scores for the various HSAP achievement levels are presented in table 6.3.

TABLE 6.3
Rasch Ability and Scale Score Cut Scores for HSAP Achievement Levels

	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.

6.8 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:

- 1 - *Needs improvement*: if $\theta < \theta_c - 1$ SE
- 2 - *May need improvement*: if $\theta_c - 1$ SE $\leq \theta < \theta_c + 1$ SE
- 3 - *[No descriptor used]* if $\theta \geq \theta_c + 1$ 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 HSAP Content-Area Classifications

Content Area	Rasch Ability (θ)	SE(θ)	68% Confidence Interval	
			$\theta - 1SE$	$\theta + 1SE$
Mathematics				
Fall 2008				
Number and Operations	0.029	0.567	-0.538	0.596
Algebra	-0.219	0.486	-0.705	0.267
Measurement and Geometry	-0.051	0.480	-0.531	0.429
Data Analysis and Probability	-0.124	0.734	-0.858	0.610
Spring 2009				
Number and Operations	0.079	0.573	-0.494	0.652
Algebra	0.018	0.489	-0.471	0.507
Measurement and Geometry	-0.052	0.477	-0.529	0.425
Data Analysis and Probability	-0.104	0.735	-0.839	0.631
Summer 2009				
Number and Operations	0.044	0.574	-0.530	0.618
Algebra	-0.037	0.478	-0.515	0.441
Measurement and Geometry	-0.056	0.476	-0.532	0.420
Data Analysis and Probability	-0.165	0.730	-0.895	0.565
ELA				
Fall 2008				
Reading Process and Comprehension	0.037	0.433	-0.396	0.470
Analysis of Texts	0.205	0.515	-0.310	0.720
Word Study and Analysis	0.317	0.770	-0.453	1.087
Research	0.460	0.812	-0.352	1.272
Writing	0.187	0.377	-0.190	0.564
Spring 2009				
Reading Process and Comprehension	0.066	0.406	-0.340	0.472
Analysis of Texts	0.122	0.540	-0.418	0.662
Word Study and Analysis	0.299	0.765	-0.466	1.064
Research	0.610	0.793	-0.183	1.403
Writing	0.161	0.417	-0.256	0.578
Summer 2009				
Reading Process and Comprehension	0.089	0.444	-0.355	0.533
Analysis of Texts	0.288	0.517	-0.229	0.805
Word Study and Analysis	0.674	0.851	-0.177	1.525
Research	0.197	0.724	-0.527	0.921
Writing	0.185	0.406	-0.221	0.591

6.9 PERCENTAGE OF STUDENTS IN EACH ACHIEVEMENT LEVEL

Tables 6.6 through 6.11, below, present student performance on the fall, spring, and summer HSAP tests for mathematics and ELA. Percentages of students in the four achievement levels are reported for all students and for various subgroups. The summary includes all students who were tested but excludes students in adult education and district-approved homeschools. Tables 6.12 through 6.17 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.

TABLE 6.5
Percentage of First Time Students Passing Both Tests

	Fall 2008	Spring 2009	Summer 2009
Percent Passed	35.97	76.23	100.00

TABLE 6.6
Fall 2008 HSAP Mathematics Operational Test: Percentage of Students
in Achievement Levels Overall and by Subgroups (All Attempts)

Subgroup	Achievement Levels				At or	At or	N
	Level 1	Level 2	Level 3	Level 4	Above	Above	
					Level	Level	
					2	3	
Overall	59.7	27	8.1	5.2	40.3	13.3	11,248
Gender							
Female	58.9	27.9	8.4	4.7	41.1	13.1	3,000
Male	61.9	26.2	6.8	5	38.1	11.8	3,552
Invalid	38	27.4	20.5	14.2	62	34.7	161
Ethnicity							
African American	71.8	24.5	2.8	0.8	28.2	3.6	4,611
Asian/Pacific Islander	25.4	20.3	28.8	25.4	74.6	54.2	30
Hispanic	57.4	30.3	7.7	4.6	42.6	12.3	350
American Indian	64.3	25	10.7	--	35.7	10.7	18
White	42.4	31.3	14.7	11.6	57.6	26.3	1,450
Other	41.9	28	19.9	10.2	58.1	30.1	103
Unknown	37.4	27	22.3	13.4	62.6	35.6	151
ESL*							
Yes	51.9	25.9	14.3	7.9	48.1	22.2	138
No	46	28.8	14.9	10.2	54	25.1	1,662
Unknown	66.7	26.2	4.5	2.6	33.3	7.2	4,913
Lunch Program							
No free/reduced lunch	36.3	29.3	18.7	15.7	63.7	34.4	986
Free lunch	64.1	26.4	7.4	2.1	35.9	9.5	1,445
Reduced lunch	49.4	24.7	17	8.9	50.6	25.9	128
Unknown	69	26.4	3.1	1.4	31	4.6	4,154
IEP**							
Yes	55.5	28.9	9.4	6.1	44.5	15.6	5,216
No	80.9	17.6	1.1	0.4	19.1	1.6	1,497
Unknown	--	--	--	--	0	0	--
Migrant							
Yes	33.3	33.3	33.3	--	66.7	33.3	1
No	49.1	27.8	13.7	9.4	50.9	23.2	2,477
Unknown	68.4	26.4	3.5	1.7	31.6	5.2	4,235

* English as a second language

** individualized education program

TABLE 6.7
Spring 2009 HSAP Mathematics Operational Test: Percentage of Students
in Achievement Levels Overall and by Subgroups (First Attempt)

Subgroup	Achievement Levels				At or	At or	N
	Level 1	Level 2	Level 3	Level 4	Above	Above	
					Level	Level	
					2	3	
Overall	20.8	29.3	24.1	25.8	79.2	50.0	53,323
Gender							
Female	18.9	30.4	25.5	25.1	81.1	50.6	26,770
Male	22.5	28.1	22.8	26.6	77.5	49.4	26,468
Invalid	41.2	30.6	14.1	14.1	58.8	28.2	85
Ethnicity							
African American	33.2	36.4	20.6	9.9	66.8	30.4	20,851
Asian/Pacific Islander	7.1	15.2	22.2	55.5	92.9	77.7	618
Hispanic	22.9	32.9	24.3	19.9	77.1	44.2	2,175
American Indian	22.7	29.1	21.8	26.4	77.3	48.2	110
White	11.9	24.0	26.8	37.3	88.1	64.1	28,686
Other	15.4	31.9	25.2	27.6	84.6	52.7	747
Unknown	41.9	31.6	15.4	11.0	58.1	26.5	136
ESL*							
Parent Waiver	19.5	28.7	26.4	25.3	80.5	51.7	87
Pre-Functional - Advanced	34.2	35.2	18.9	11.8	65.8	30.6	1,156
Initially English Proficient	5.9	19.1	32.4	42.6	94.1	75.0	68
Title III Exited	6.5	29.0	30.0	34.5	93.5	64.5	200
English Speaker I	12.0	20.8	24.5	42.7	88.0	67.2	192
English Speaker II	20.1	29.2	24.4	26.3	79.9	50.7	50,669
All others	45.0	27.3	15.2	12.4	55.0	27.7	951
Lunch Program							
No free/reduced lunch	12.2	24.0	26.3	37.5	87.8	63.8	28,624
Free lunch	32.4	35.7	20.8	11.0	67.6	31.8	20,921
Reduced lunch	20.5	33.5	26.3	19.6	79.5	46.0	3,778
IEP**							
Yes	64.7	23.8	8.2	3.3	35.3	11.5	6,381
No	14.8	30.0	26.3	28.9	85.2	55.2	46,942
Migrant							
Yes	14.3	42.9	28.6	14.3	85.7	42.9	14
No	20.8	29.3	24.1	25.8	79.2	50.0	53,309

* English as a second language

** individualized education program

TABLE 6.8

Summer 2009 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts)

Subgroup	Achievement Levels				At or Above Level 2	At or Above Level 3	N
	Level 1	Level 2	Level 3	Level 4			
Overall	63.1	33.0	3.9	--	36.9	3.9	203
Gender							
Female	69.7	29.4	0.9	--	30.3	0.9	109
Male	59.2	36.8	3.9	--	40.8	3.9	76
Invalid	38.9	38.9	22.2	--	61.1	22.2	18
Ethnicity							
African American	67.1	30.9	2.0	--	32.9	2.0	152
Hispanic	50.0	50.0	--	--	50.0	--	2
American Indian	--	100.0	--	--	100.0	--	1
White	59.1	36.4	4.5	--	40.9	4.5	22
Other	66.7	33.3	--	--	33.3	--	3
Unknown	43.5	39.1	17.4	--	56.5	17.4	23
ESL*							
Pre-Functional - Advanced	50.0	50.0	0.0	0.0	50.0	0.0	2
English Speaker II	61.6	37.0	1.4	0.0	38.4	1.4	73
All others	64.1	30.5	5.5	0.0	35.9	5.5	128
Lunch Program							
No free/reduced lunch	59.2	36.2	4.6	--	40.8	4.6	130
Free lunch	71.6	25.4	3.0	--	28.4	3.0	67
Reduced lunch	50.0	50.0	--	--	50.0	--	6
IEP**							
Yes	65.2	32.6	2.2	--	34.8	2.2	46
No	62.4	33.1	4.5	--	37.6	4.5	157
Migrant							
Yes	--	--	--	--	--	--	0
No	63.1	33.0	3.9	--	36.9	3.9	203

* English as a second language
 ** individualized education program

TABLE 6.9
Fall 2008 HSAP ELA Operational Test: Percentage of Students
in Achievement Levels Overall and by Subgroups (All Attempts)

Subgroup	Achievement Levels				At or	At or	N
	Level 1	Level 2	Level 3	Level 4	Above	Above	
					Level	Level	
					2	3	
Overall	47.3	30.2	10.6	11.9	52.7	22.5	8,886
Gender							
Female	42.1	29.5	12.3	16.1	57.9	28.4	1,484
Male	51.1	30.9	9.4	8.6	48.9	18	2,643
Invalid	38.5	25.5	14.6	21.4	61.5	35.9	74
Ethnicity							
African American	58.9	31.9	6.2	3	41.1	9.2	2,747
Asian/Pacific Islander	37.7	29.2	16.2	16.9	62.3	33.1	58
Hispanic	56.6	28.6	8.5	6.3	43.4	14.8	368
American Indian	50	31.8	9.1	9.1	50	18.2	11
White	29.6	28.2	16.6	25.7	70.4	42.2	884
Other	28.4	31.4	23.5	16.7	71.6	40.2	58
Unknown	36.1	26.9	14.4	22.6	63.9	37	75
ESL*							
Yes	56.6	32	6.6	4.7	43.4	11.4	179
No	32.2	27.7	18.6	21.5	67.8	40.1	1,077
Unknown	56.4	31.7	5.8	6.1	43.6	11.9	2,945
Lunch Program							
No free/reduced lunch	22.4	27.1	20	30.5	77.6	50.4	581
Free lunch	51.6	31.6	11	5.7	48.4	16.8	1,051
Reduced lunch	36.9	26.1	18.5	18.5	63.1	36.9	82
Unknown	61.6	31.7	4	2.7	38.4	6.7	2,487
IEP**							
Yes	42.4	31.4	12.1	14.1	57.6	26.3	3,124
No	71.3	24.5	3.2	1	28.7	4.2	1,077
Unknown	--	--	--	--	0	0	--
Migrant							
Yes	55.6	22.2	11.1	11.1	44.4	22.2	5
No	36	28.6	16	19.3	64	35.4	1,672
Unknown	59.7	32	4.7	3.7	40.3	8.3	2,524

* English as a second language

**individualized education program

TABLE 6.10
Spring 2009 HSAP ELA Operational Test: Percentage of Students
in Achievement Levels Overall and by Subgroups (First Attempt)

Subgroup	Achievement Levels				At or	At or	N
	Level 1	Level 2	Level 3	Level 4	Above	Above	
					Level	Level	
					2	3	
Overall	15.7	35.0	28.4	21.0	84.3	49.3	53,088
Gender							
Female	11.8	34.4	30.0	23.8	88.2	53.8	26,604
Male	19.7	35.5	26.7	18.1	80.3	44.8	26,409
Invalid	33.3	34.7	14.7	17.3	66.7	32.0	75
Ethnicity							
African American	24.2	45.3	22.1	8.4	75.8	30.5	20,679
Asian/Pacific Islander	12.2	24.8	28.9	34.1	87.8	63.0	616
Hispanic	24.5	37.8	25.9	11.8	75.5	37.7	2,161
American Indian	18.9	38.7	21.6	20.7	81.1	42.3	111
White	9.0	27.5	33.1	30.4	91.0	63.5	28,662
Other	11.6	34.4	31.1	22.9	88.4	54.0	742
Unknown	33.3	38.5	17.9	10.3	66.7	28.2	117
ESL*							
Parent Waiver	17.2	49.4	29.9	3.4	82.8	33.3	87
Pre-Functional - Advanced	41.7	40.5	14.4	3.3	58.3	17.7	1,145
Initially English Proficient	2.9	28.6	38.6	30.0	97.1	68.6	70
Title III Exited	5.0	35.3	41.3	18.4	95.0	59.7	201
English Speaker I	9.9	28.1	31.3	30.7	90.1	62.0	192
English Speaker II	14.9	34.8	28.8	21.5	85.1	50.3	50,485
All others	34.1	38.0	16.9	11.0	65.1	27.9	908
Lunch Program							
No free/reduced lunch	8.4	27.4	32.9	31.4	91.6	64.3	28,507
Free lunch	26.0	44.2	22.0	7.8	74.0	29.8	20,825
Reduced lunch	14.6	41.1	29.4	14.9	85.4	44.3	3,756
IEP**							
Yes	58.1	33.0	7.3	1.6	41.9	8.9	6,386
No	9.9	35.2	31.2	23.6	90.1	54.8	46,702
Migrant							
Yes	46.2	30.8	23.1	0.0	53.8	23.1	13
No	15.7	35.0	28.4	21.0	84.3	49.3	53,075

* English as a second language

** individualized education program

TABLE 6.11
Summer 2009 HSAP ELA Operational Test: Percentage of Students
in Achievement Levels Overall and by Subgroups (All Attempts)

Subgroup	Achievement Levels				At or	At or	N
	Level 1	Level 2	Level 3	Level 4	Above	Above	
					Level	Level	
					2	3	
Overall	48.8	46.3	3.7	1.2	51.2	4.9	82
Gender							
Female	45.0	50.0	2.5	2.5	55.0	5.0	40
Male	50.0	47.4	2.6	--	50.0	2.6	38
Invalid	75.0	--	25.0	--	25.0	25.0	4
Ethnicity							
African American	48.3	48.3	3.4	--	51.7	3.4	58
Asian/Pacific Islander	100.0	--	--	--	--	--	3
Hispanic	40.0	60.0	--	--	60.0	--	5
White	33.3	55.6	--	11.1	66.7	11.1	9
Other	100.0	--	--	--	--	--	1
Unknown	50.0	33.3	16.7	--	50.0	16.7	6
ESL*							
Pre-Functional - Advanced	66.7	33.3	--	--	33.3	--	6
English Speaker II	36.4	54.5	6.1	3.0	63.6	9.1	33
All others	55.8	41.9	2.3	--	44.2	2.3	43
Lunch Program							
No free/reduced lunch	46.7	48.9	2.2	2.2	53.3	4.4	45
Free lunch	55.9	41.2	2.9	--	44.1	2.9	34
Reduced lunch	--	66.7	33.3	--	100.0	33.3	3
IEP**							
Yes	40.9	50.0	9.1	--	59.1	9.1	22
No	51.7	45.0	1.7	1.7	48.3	3.3	60
Migrant							
Yes	--	--	--	--	--	--	0
No	48.8	46.3	3.7	1.2	51.2	4.9	82

* English as a second language
** individualized education program

TABLE 6.12
Fall 2008 HSAP Mathematics Operational Test:
Content-Area Information (All Attempts)

Subgroup	Level 1				Level 2 and Above			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	N1*	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	N2**
Number and Operations								
All students	73.0%	25.4%	1.6%	6,713	8.1%	39.0%	52.9%	4,535
Females	72.9%	25.8%	1.3%	3,000	9.9%	40.4%	49.7%	2,090
Males	72.7%	25.4%	1.9%	3,552	6.8%	39.9%	53.3%	2,182
African Americans	73.9%	24.9%	1.2%	4,611	11.8%	52.6%	35.6%	1,808
Whites	69.8%	27.2%	3.0%	1,450	5.2%	30.6%	64.2%	1,973
Algebra								
All students	32.5%	64.2%	3.2%	6,713	1.4%	49.5%	49.1%	4,535
Females	27.5%	68.8%	3.8%	3,000	1.1%	50.1%	48.8%	2,090
Males	36.7%	60.5%	2.8%	3,552	1.7%	50.5%	47.8%	2,182
African Americans	30.8%	65.8%	3.4%	4,611	1.5%	61.7%	36.8%	1,808
Whites	37.4%	60.1%	2.5%	1,450	1.5%	42.8%	55.7%	1,973
Measurement and Geometry								
All students	67.4%	32.4%	0.2%	6,713	10.0%	53.5%	36.5%	4,535
Females	69.4%	30.4%	0.1%	3,000	10.8%	55.6%	33.6%	2,090
Males	65.7%	34.0%	0.3%	3,552	10.1%	53.3%	36.6%	2,182
African Americans	69.8%	30.1%	0.1%	4,611	16.1%	67.6%	16.3%	1,808
Whites	60.1%	39.3%	0.6%	1,450	5.9%	44.6%	49.5%	1,973
Data Analysis and Probability								
All students	30.7%	65.3%	4.0%	6,713	2.1%	48.1%	49.9%	4,535
Females	26.3%	68.6%	5.1%	3,000	1.7%	48.6%	49.7%	2,090
Males	34.1%	62.8%	3.1%	3,552	2.6%	49.4%	48.0%	2,182
African Americans	30.4%	65.7%	3.8%	4,611	3.0%	59.6%	37.4%	1,808
Whites	30.3%	65.2%	4.5%	1,450	1.5%	40.3%	58.2%	1,973

* total number of students in Level 1

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

TABLE 6.13
Spring 2009 HSAP Mathematics Operational Test:
Content-Area Information (First Attempt)

Subgroup	Level 1				Level 2 and Above			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<i>N1*</i>	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<i>N2**</i>
Number and Operations								
All students	79.5%	19.5%	1.0%	11,070	6.2%	28.2%	65.6%	42,253
Females	80.5%	18.7%	0.8%	5,071	7.2%	29.3%	63.5%	21,699
Males	78.6%	20.2%	1.2%	5,964	5.3%	27.0%	67.7%	20,504
African Americans	80.4%	18.7%	0.9%	6,919	9.8%	38.5%	51.7%	13,932
Whites	77.5%	21.1%	1.3%	3,412	4.2%	22.5%	73.3%	25,274
Algebra								
All students	42.4%	55.1%	2.4%	11,070	0.6%	24.0%	75.4%	42,253
Females	37.4%	59.7%	2.9%	5,071	0.6%	23.2%	76.2%	21,699
Males	46.7%	51.3%	2.1%	5,964	0.7%	24.8%	74.6%	20,504
African Americans	41.8%	55.6%	2.6%	6,919	0.7%	31.5%	67.8%	13,932
Whites	44.3%	53.9%	1.8%	3,412	0.6%	20.1%	79.3%	25,274
Measurement and Geometry								
All students	61.5%	38.1%	0.4%	11,070	2.3%	30.3%	67.3%	42,253
Females	63.4%	36.4%	0.2%	5,071	2.6%	32.1%	65.2%	21,699
Males	59.8%	39.5%	0.7%	5,964	2.0%	28.4%	69.6%	20,504
African Americans	62.8%	36.8%	0.3%	6,919	4.1%	45.9%	50.0%	13,932
Whites	57.5%	41.8%	0.6%	3,412	1.4%	21.6%	77.0%	25,274
Data Analysis and Probability								
All students	30.8%	60.4%	8.8%	11,070	0.8%	21.8%	77.5%	42,253
Females	25.2%	64.2%	10.6%	5,071	0.5%	21.0%	78.5%	21,699
Males	35.6%	57.2%	7.3%	5,964	1.0%	22.6%	76.4%	20,504
African Americans	30.8%	60.7%	8.5%	6,919	0.9%	32.2%	66.9%	13,932
Whites	30.6%	60.0%	9.4%	3,412	0.6%	15.9%	83.4%	25,274

* total number of students in Level 1

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

TABLE 6.14
Summer 2009 HSAP Mathematics Operational Test:
Content-Area Information (All Attempts)

Subgroup	Level 1				Level 2 and Above			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<i>N1*</i>	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<i>N2**</i>
Number and Operations								
All students	66.4%	32.0%	1.6%	128	14.7%	53.3%	32.0%	75
Females	68.4%	30.3%	1.3%	76	12.1%	63.6%	24.2%	33
Males	57.8%	40.0%	2.2%	45	19.4%	54.8%	25.8%	31
African Americans	67.6%	31.4%	1.0%	102	12.0%	62.0%	26.0%	50
Whites	46.2%	46.2%	7.7%	13	44.4%	33.3%	22.2%	9
Algebra								
All students	62.5%	35.9%	1.6%	128	6.7%	66.7%	26.7%	75
Females	65.8%	31.6%	2.6%	76	6.1%	69.7%	24.2%	33
Males	57.8%	42.2%	--	45	9.7%	61.3%	29.0%	31
African Americans	61.8%	37.3%	1.0%	102	4.0%	66.0%	30.0%	50
Whites	61.5%	30.8%	7.7%	13	22.2%	55.6%	22.2%	9
Measurement and Geometry								
All students	64.8%	34.4%	0.8%	128	10.7%	80.0%	9.3%	75
Females	65.8%	34.2%	--	76	12.1%	78.8%	9.1%	33
Males	68.9%	28.9%	2.2%	45	9.7%	83.9%	6.5%	31
African Americans	66.7%	32.4%	1.0%	102	12.0%	80.0%	8.0%	50
Whites	53.8%	46.2%	--	13	0.0%	88.9%	11.1%	9
Data Analysis and Probability								
All students	7.0%	73.4%	19.5%	128	--	46.7%	53.3%	75
Females	6.6%	72.4%	21.1%	76	--	42.4%	57.6%	33
Males	8.9%	75.6%	15.6%	45	--	54.8%	45.2%	31
African Americans	7.8%	69.6%	22.5%	102	--	52.0%	48.0%	50
Whites	--	100.0%	--	13	--	11.1%	88.9%	9

* total number of students in Level 1

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

TABLE 6.15
Fall 2008 HSAP ELA Operational Test:
Content-Area Information (All Attempts)

Subgroup	Level 1				Level 2 and Above			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	N1*	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	N2**
Reading Process and Comprehension								
All students	46.1%	49.4%	4.5%	4,201	1.8%	28.7%	69.5%	4,685
Females	43.6%	52.6%	3.8%	1,484	1.3%	25.2%	73.5%	2,039
Males	47.4%	47.9%	4.7%	2,643	2.2%	31.8%	66.0%	2,528
African Americans	46.6%	49.2%	4.2%	2,747	2.8%	39.9%	57.3%	1,913
Whites	44.8%	49.9%	5.3%	884	1.1%	19.3%	79.6%	2,104
Analysis of Texts								
All students	61.1%	34.1%	4.8%	4,201	9.4%	34.0%	56.6%	4,685
Females	61.7%	35.0%	3.4%	1,484	9.8%	33.3%	56.9%	2,039
Males	60.8%	33.5%	5.6%	2,643	9.3%	34.8%	56.0%	2,528
African Americans	63.0%	32.8%	4.2%	2,747	15.5%	46.6%	37.9%	1,913
Whites	57.9%	35.5%	6.6%	884	4.3%	23.2%	72.5%	2,104
Word Study and Analysis								
All students	46.2%	51.7%	2.1%	4,201	10.1%	53.6%	36.3%	4,685
Females	47.8%	50.1%	2.2%	1,484	10.2%	49.1%	40.7%	2,039
Males	45.2%	52.7%	2.0%	2,643	10.1%	57.6%	32.2%	2,528
African American	46.9%	51.8%	1.3%	2,747	14.4%	64.5%	21.1%	1,913
White	43.4%	53.2%	3.4%	884	6.7%	42.6%	50.8%	2,104
Writing								
All students	42.9%	53.7%	3.3%	4,201	7.1%	54.9%	38.0%	4,685
Females	39.6%	56.1%	4.4%	1,484	6.6%	51.7%	41.7%	2,039
Males	44.9%	52.3%	2.8%	2,643	7.7%	57.6%	34.7%	2,528
African Americans	41.8%	54.7%	3.5%	2,747	9.0%	62.9%	28.1%	1,913
Whites	48.0%	49.3%	2.7%	884	5.9%	47.7%	46.4%	2,104
Research								
All students	84.6%	14.8%	0.7%	4,201	11.9%	31.5%	56.5%	4,685
Females	83.3%	15.9%	0.8%	1,484	10.1%	27.5%	62.4%	2,039
Males	85.3%	14.2%	0.5%	2,643	13.7%	35.1%	51.2%	2,528
African Americans	84.7%	14.7%	0.6%	2,747	16.3%	42.0%	41.7%	1,913
Whites	80.8%	18.3%	0.9%	884	6.9%	24.0%	69.1%	2,104

* total number of students in Level 1

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

TABLE 6.16
Spring 2009 HSAP ELA Operational Test:
Content-Area Information (First Attempt)

Subgroup	Level 1				Level 2 and Above			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<i>N1*</i>	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<i>N2**</i>
Reading Process and Comprehension								
All students	55.6%	39.5%	4.8%	8,354	0.6%	14.9%	84.4%	44,734
Females	50.9%	43.8%	5.3%	3,134	0.5%	13.2%	86.3%	23,470
Males	58.5%	37.0%	4.5%	5,195	0.8%	16.8%	82.4%	21,214
African Americans	55.4%	39.8%	4.7%	5,013	0.8%	21.5%	77.7%	15,666
Whites	55.1%	39.7%	5.2%	2,590	0.5%	11.0%	88.5%	26,072
Analysis of Texts								
All students	67.2%	27.0%	5.7%	8,354	2.6%	19.5%	77.9%	44,734
Females	66.3%	28.7%	5.0%	3,134	2.5%	19.4%	78.1%	23,470
Males	67.8%	26.1%	6.1%	5,195	2.7%	19.6%	77.7%	21,214
African Americans	69.0%	26.1%	4.9%	5,013	4.3%	28.9%	66.8%	15,666
Whites	64.2%	28.6%	7.3%	2,590	1.5%	13.9%	84.6%	26,072
Word Study and Analysis								
All students	48.3%	48.7%	3.0%	8,354	3.0%	52.3%	44.7%	44,734
Females	47.7%	49.9%	2.4%	3,134	3.3%	55.9%	40.8%	23,470
Males	48.8%	47.9%	3.4%	5,195	2.7%	48.4%	48.9%	21,214
African Americans	48.9%	48.8%	2.3%	5,013	4.7%	63.9%	31.4%	15,666
Whites	46.1%	49.3%	4.5%	2,590	1.9%	45.1%	52.9%	26,072
Writing								
All students	78.8%	19.8%	1.4%	8,354	5.0%	25.4%	69.7%	44,734
Females	75.5%	22.8%	1.8%	3,134	3.8%	22.8%	73.4%	23,470
Males	80.8%	18.0%	1.3%	5,195	6.3%	28.1%	65.6%	21,214
African Americans	78.9%	19.9%	1.2%	5,013	7.4%	34.2%	58.4%	15,666
Whites	77.8%	20.1%	2.0%	2,590	3.4%	19.9%	76.7%	26,072
Research								
All students	50.2%	48.0%	1.8%	8,354	8.4%	56.2%	35.5%	44,734
Females	51.3%	47.2%	1.5%	3,134	8.8%	56.0%	35.1%	23,470
Males	49.7%	48.4%	1.9%	5,195	7.9%	56.3%	35.8%	21,214
African Americans	51.2%	47.3%	1.5%	5,013	13.3%	65.2%	21.5%	15,666
Whites	49.3%	48.2%	2.5%	2,590	5.5%	50.4%	44.2%	26,072

* total number of students in Level 1

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

TABLE 6.17
Summer 2009 HSAP ELA Operational Test:
Content-Area Information (All Attempts)

Subgroup	Level 1				Level 2 and Above			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	N1*	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	N2**
Reading Process and Comprehension								
All students	32.5%	67.5%	--	40	--	47.6%	52.4%	42
Females	44.4%	55.6%	--	18	--	54.5%	45.5%	22
Males	26.3%	73.7%	--	19	--	42.1%	57.9%	19
African Americans	39.3%	60.7%	--	28	--	50.0%	50.0%	30
Whites	--	100.0%	--	3	--	50.0%	50.0%	6
Analysis of Texts								
All students	37.5%	62.5%	--	40	--	50.0%	50.0%	42
Females	55.6%	44.4%	--	18	--	63.6%	36.4%	22
Males	21.1%	78.9%	--	19	--	36.8%	63.2%	19
African Americans	35.7%	64.3%	--	28	--	46.7%	53.3%	30
Whites	--	100.0%	--	3	--	50.0%	50.0%	6
Word Study and Analysis								
All students	62.5%	37.5%	--	40	23.8%	59.5%	16.7%	42
Females	72.2%	27.8%	--	18	31.8%	50.0%	18.2%	22
Males	57.9%	42.1%	--	19	15.8%	73.7%	10.5%	19
African American	64.3%	35.7%	--	28	20.0%	63.3%	16.7%	30
White	66.7%	33.3%	--	3	16.7%	66.7%	16.7%	6
Writing								
All students	50.0%	45.0%	5.0%	40	2.4%	47.6%	50.0%	42
Females	50.0%	38.9%	11.1%	18	4.5%	31.8%	63.6%	22
Males	52.6%	47.4%	--	19	--	68.4%	31.6%	19
African Americans	53.6%	42.9%	3.6%	28	3.3%	53.3%	43.3%	30
Whites	33.3%	66.7%	--	3	--	50.0%	50.0%	6
Research								
All students	37.5%	60.0%	2.5%	40	19.0%	66.7%	14.3%	42
Females	44.4%	50.0%	5.6%	18	18.2%	63.6%	18.2%	22
Males	36.8%	63.2%	--	19	21.1%	73.7%	5.3%	19
African Americans	42.9%	53.6%	3.6%	28	20.0%	70.0%	10.0%	30
Whites	33.3%	66.7%	--	3	16.7%	66.7%	16.7%	6

* total number of students in Level 1

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

6.10 DESCRIPTIVE STATISTICS

Descriptive statistics for scale score distributions are presented in table 6.18 for students overall and by gender and selected ethnic group.

TABLE 6.18
HSAP Summary Statistics Overall and by Subgroups (All Attempts)

Subgroup	Mathematics			ELA		
	N	Scale Score		N	Scale Score	
		Mean	Std. Dev.		Mean	Std. Dev.
Fall 2008						
All students	11,248	199.5	20.9	8,886	205.3	26.1
Females	5,090	200.0	19.6	3,523	209.5	27.0
Males	5,734	198.1	21.2	5,171	202.2	24.8
African Americans	6,419	193.2	14.2	4,660	196.9	19.0
Whites	3,423	208.3	25.2	2,988	218.4	29.8
Spring 2009						
All students	58,592	220.1	29.9	56,428	220.1	24.4
Females	29,234	220.7	28.5	27,877	223.3	23.3
Males	29,216	219.7	31.3	28,453	216.9	25.0
African Americans	24,531	206.9	23.4	22,837	210.3	21.9
Whites	29,838	230.9	30.3	29,397	228.3	23.1
Summer 2009						
All students	203	195.8	12.9	82	198.5	14.2
Females	109	194.3	10.1	40	199.1	15.3
Males	76	195.8	15.0	38	197.6	12.3
African Americans	152	194.6	12.7	58	198.0	13.2
Whites	22	198.5	10.1	9	205.1	15.8

Chapter 7

RELIABILITY

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

7.1 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

$$\text{strat } \alpha \rho_{XX'} = 1 - \frac{\sum \sigma_{Y_j}^2 (1 - \alpha \rho_{Y_j Y_j'})}{\sigma_X^2}, \text{ where } \sigma_X^2 = \text{the total score variance; } \sigma_{Y_j}^2 = \text{the score variance}$$

for a part-test j ; and $\alpha \rho_{Y_j Y_j'}$ = the reliability of the part-test j .

Table 7.1 presents the reliability coefficients and SEM for mathematics and ELA for all students and subgroups. The maximum possible raw score is 71 in mathematics and 96 in ELA.

TABLE 7.1
Reliability Coefficients and SEM for HSAP Raw Scores

	Fall 2008 (All Attempts)	Spring 2009 (First Attempts)	Summer 2009 (All Attempts)
Mathematics			
Reliability	0.91	0.95	0.79
SEM	3.77	3.26	3.65
ELA			
Reliability	0.95	0.95	0.87
SEM	4.82	3.48	4.02

7.2 OVERALL AND CONDITIONAL SEM

Table 7.2 presents the classical test-theory SEM and the IRT-based conditional SEM at the scale score cut points. The SEM in the table are reported in units of scale score 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 cut 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 cut 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 7.2
Classical and Conditional SEM for HSAP

Subject	Classical SEM	IRT-Based Conditional SEM		
		Level 2	Level 3	Level 4
Mathematics, Fall 2008	6.47	5.41	5.88	7.59
Mathematics, Spring 2009	6.80	5.69	6.19	7.78
Mathematics, Summer 2009	5.97	5.60	6.01	7.83
ELA, Fall 2008	7.53	5.16	5.81	7.11
ELA, Spring 2009	5.54	5.45	6.41	7.82
ELA, Summer 2009	5.02	5.45	6.43	7.89

Note: Spring statistics include only students taking the test for the first time.

7.3 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, APA, and NCME 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), this report uses the beta binomial method (Huynh 1979). Table 7.3 presents a summary of agreements between the operational test classifications—that is, the percentages of students who would be consistently classified in the same achievement levels on two equivalent administrations of the test.

TABLE 7.3
Consistency Indexes for HSAP Achievement Levels

	Level 2	Level 3
Fall 2008 (All Attempts)		
Mathematics	85.6	92.2
ELA	88.6	90.7
Spring 2009 (First Attempts)		
Mathematics	92.4	89.1
ELA	92.9	88.1
Summer 2009 (All Attempts)		
Mathematics	77.2	96.1
ELA	81.0	95.2

Chapter 8

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 HSAP test items with reference to the state academic standards. Evidence on item fairness is examined with the information on differential item functioning. Evidence on internal structure is provided in correlations among content areas.

8.1 ITEM DISTRIBUTION ACROSS STRANDS

The HSAP test forms were constructed from precalibrated item pools that had been created on the basis of the 2003 census field-test results. An analysis of field-test statistics determined that all items in these pools adequately measured specific knowledge and skills deemed appropriate for assessment by standardized tests. All items were reviewed by the Content Review Committee and the Sensitivity Review Committee (SRC) and approved by the SCDE. The HSAP test specifications are presented in section 4.2, above, in terms of distribution of score point values by content area.

8.2 ITEM DEVELOPMENT

All HSAP items were developed in alignment with the South Carolina academic and measurement guidelines. Various committees reviewed all items; only items reviewed by these committees and approved by the SCDE were included in the operational forms.

8.3 DIFFERENTIAL ITEM FUNCTIONING (DIF)

An important goal of test development is establishing an item pool that is fair to all students. All HSAP items were therefore reviewed for potential bias and for DIF. Specifically, the SRC reviewed the test items for bias with regard to language that might disadvantage a particular group of students, might be considered offensive to members of a particular group, or might present obstacles to a certain group due to factors unrelated to the content and processes specified in the state academic standards.

After data were collected, the DIF statistics were produced for the statistical review. A psychometric definition of the term *test fairness* is the degree to which an item performs similarly for different groups 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 not necessarily be seen 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 were used by Pearson 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). DRC 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 the SMD divided by the 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 the SMD divided by the standard deviation of the item score (i.e., $|SMD/SD|$) is larger than .17.
- All other items are classified as A category.

The Polytomous (Constructed-Response) DIF classification was defined as:

- Rule 1: $|Z_{SMD}| < 2.00$ was classified as A.
- Rule 2: $2.00 \leq |Z_{SMD}| < 5.00$ was classified as B.

- Rule 3: $5.00 \leq |Z_{SMD}|$ was classified as C.

When items for the operational forms were selected, each item’s statistics from the initial field test were reviewed and approved by the SCDE. The inclusion of any “flagged” items on an operational form (i.e., items classified as C category) was possible only when the SCDE approved such inclusion.

Examining item results for DIF requires the use of a statistical test. When applied to large numbers of items, it is to be expected that a few items might be classified as category C due to Type I (false positive) errors. SCDE staff examined every flagged field test item for any potential sources of DIF. If none was found, the item was deemed acceptable for use on an operational form. Items on an operational form may be flagged for the same reason. Items that continue to be flagged for DIF are removed from the item bank. For the fall 2008, spring 2009, and summer 2009 operational forms, Tables 8.1, 8.2, and 8.3 report the numbers of items in the various DIF categories for ELA and mathematics.

TABLE 8.1
Fall 2008 Summary of DIF Classifications for HSAP Mathematics and ELA Items

Item Type	Reference Group	Focal Group	Total N of Items	DIF Classification		
				A	B	C
Mathematics						
Multiple choice	Male	Female	62	58	3	1
Multiple choice	White	Black	62	53	6	3
Constructed response	Male	Female	3	3	0	0
Constructed response	White	Black	3	3	0	0
ELA						
Multiple choice	Male	Female	58	56	2	0
Multiple choice	White	Black	58	56	2	0
Constructed response	Male	Female	2	2	0	0
Constructed response	White	Black	2	2	0	0
Extended response	Male	Female	1	1	0	0
Extended response	White	Black	1	1	0	0

TABLE 8.2
Spring 2009 Summary of DIF Classifications for HSAP Mathematics and ELA Items

Item Type	Reference Group	Focal Group	Total N of Items	DIF Classification		
				A	B	C
Mathematics						
Multiple choice	Male	Female	62	55	7	0
Multiple choice	White	Black	62	59	3	0
Constructed response	Male	Female	3	0	0	3
Constructed response	White	Black	3	0	0	3
ELA						
Multiple choice	Male	Female	60	57	2	1
Multiple choice	White	Black	60	57	3	0
Constructed response	Male	Female	2	0	0	2
Constructed response	White	Black	2	0	0	2
Extended response	Male	Female	4	0	3	1
Extended response	White	Black	4	0	0	4

TABLE 8.3
Summer 2009 Summary of DIF Classifications for HSAP Mathematics and ELA Items

Item Type	Reference Group	Focal Group	Total N of Items	DIF Classification*		
				A	B	C
Mathematics						
Multiple choice	Male	Female	62	47	7	8
Multiple choice	White	Black	62	30	17	15
Constructed response	Male	Female	3	3	0	0
Constructed response	White	Black	3	3	0	0
ELA						
Multiple choice	Male	Female	60	32	9	19
Multiple choice	White	Black	60	17	11	30
Constructed response	Male	Female	2	2	0	0
Constructed response	White	Black	2	0	0	0
Extended response	Male	Female	4	0	0	0
Extended response	White	Black	4	0	0	0

*DIF could not be calculated on all items due to identical scores in the focal or reference group.

8.4 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 8.4, 8.5, and 8.6 report the correlation matrices among the reporting category scores.

TABLE 8.4
Fall 2008 HSAP Correlations among Reporting Categories (All Attempts)

Mathematics (N=11,248)						ELA (N=8,886)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
NO	1	0.64	0.65	0.61	0.63	RC	1	0.71	0.61	0.61	0.70
AL	--	1	0.62	0.56	0.60	AT	--	1	0.60	0.57	0.61
MG	--	--	1	0.60	0.66	WS	--	--	1	0.47	0.53
DP	--	--	--	1	0.61	WR	--	--	--	1	0.53
IR	--	--	--	--	1	RS	--	--	--	--	1

TABLE 8.5
Spring 2009 HSAP Correlations among Reporting Categories (All Attempts)

Mathematics (N=58,592)						ELA (N=56,428)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
NO	1.00	0.78	0.78	0.71	0.77	RC	1.00	0.76	0.66	0.72	0.61
AL	--	1.00	0.78	0.71	0.75	AT	--	1.00	0.65	0.66	0.57
MG	--	--	1.00	0.72	0.77	WS	--	--	1.00	0.58	0.52
DP	--	--	--	1.00	0.70	WR	--	--	--	1.00	0.53
IR	--	--	--	--	1.00	RS	--	--	--	--	1.00

TABLE 8.6
Summer 2009 HSAP Correlations among Reporting Categories (All Attempts)

Mathematics (N=203)						ELA (N=82)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
NO	1.00	0.49	0.35	0.37	0.46	RC	1.00	0.76	0.44	0.51	0.29
AL	--	1.00	0.40	0.31	0.43	AT	--	1.00	0.50	0.47	0.28
MG	--	--	1.00	0.32	0.41	WS	--	--	1.00	0.25	0.17
DP	--	--	--	1.00	0.33	WR	--	--	--	1.00	0.18
IR	--	--	--	--	1.00	RS	--	--	--	--	1.00

REFERENCES

- AERA, APA, and NCME. 1999. *Standards for Educational and Psychological Testing*. 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.
- Huynh, Huynh. 1979. "Computational and Statistical Inference for Two Reliability Indices Based on the Beta-Binomial Model." *Journal of Educational Statistics* 4:231-46.
- 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.
- SCDE. 2008a. *Fall 2008 HSAP Test Administration Manual*. Columbia: South Carolina Department of Education.

- . 2008b. *Fall 2008 HSAP District Test Coordinator's Supplement*. Columbia: South Carolina Department of Education.
- . 2009a. *Spring 2009 HSAP Test Administration Manual*. Columbia: South Carolina Department of Education.
- . 2009b. *Spring 2009 HSAP District Test Coordinator's Supplement*. Columbia: South Carolina Department of Education.
- . 2009c. *Summer 2009 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.

The South Carolina Department of Education does not discriminate on the basis of race, color, national origin, sex, or disability in admission to, treatment in, or employment in its programs and activities. Inquiries regarding the nondiscrimination policies should be made to the director of the Office of Human Resources, 1429 Senate Street, Columbia, South Carolina 29201, 803-734-8505.