

**South Carolina
High School Assessment Program**

**English Language Arts and Mathematics
2007–08 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 Curriculum Services and Assessment**

**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
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 2008	2
1.2	Number of Items.....	2
2.1	Fall 2007 Summary of Student Demographics in the HSAP Sample (All Attempts).....	4
2.2	Spring 2008 Summary of Student Demographics in the HSAP Sample (All Attempts)	5
2.3	Summer 2008 Summary of Student Demographics in the HSAP Sample (All Attempts)...	6
2.4	Percentages of Students with Accommodations in the Fall 2007 HSAP Administration (All Attempts).....	8
2.5	Percentages of Students with Accommodations in the Spring 2008 HSAP Administration (All Attempts)	8
2.6	Percentages of Students with Accommodations in the Summer 2008 HSAP Administration (All Attempts)	9
2.7	Percentages of Students with Modifications in the 2007–08 HSAP Administrations (All Attempts)	10
2.8	Number of Questions in the Student Questionnaire.....	10
3.1	2007–08 HSAP Test Administration Schedule.....	11
3.2	Percentage of Students by Test Duration: HSAP Fall 2007, Spring 2008, and Summer 2008 (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 2007 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	32
6.7	Spring 2008 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (First Attempt).....	33
6.8	Summer 2008 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	34
6.9	Fall 2007 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	35
6.10	Spring 2008 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (First Attempt).....	36
6.11	Summer 2008 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	37
6.12	Fall 2007 HSAP Mathematics Operational Test: Content-Area Information (All Attempts).....	38
6.13	Spring 2008 HSAP Mathematics Operational Test: Content-Area Information (First Attempt).....	39
6.14	Summer 2008 HSAP Mathematics Operational Test: Content-Area Information (All Attempts).....	40
6.15	Fall 2007 HSAP ELA Operational Test: Content-Area Information (All Attempts).....	41
6.16	Spring 2008 HSAP ELA Operational Test: Content-Area Information (First Attempt).....	42
6.17	Summer 2008 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 2007 Summary of DIF Classifications for HSAP Mathematics and ELA Items.....	49
8.2	Spring 2008 Summary of DIF Classifications for HSAP Mathematics and ELA Items ...	49
8.3	Summer 2008 Summary of DIF Classifications for HSAP Mathematics and ELA Items.	50
8.4	Fall 2007 HSAP Correlations among Reporting Categories.....	50
8.5	Spring 2008 HSAP Correlations among Reporting Categories.....	51
8.6	Summer 2008 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 (NCLBA) of 2001 mandates that all states assess their public high school students' academic achievement in reading, language arts, and mathematics. The High School Assessment Program (HSAP) tests were developed to meet both of these statutory requirements by serving both as a criterion for a student's eligibility to receive a South Carolina high school diploma and as a primary source for reporting the required NCLBA data.

The HSAP tests were field-tested in spring 2003 to produce a sufficient number of items to build pre-equated operational test forms for both mathematics and English language arts (ELA). The first operational test was administered in spring 2004; the second and third operational tests were administered in fall 2004 and spring 2005, respectively. The first summer operational tests were administered in 2006.

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 the 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 2008

Number of Students	
ELA	881
Math	1,783

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).

TABLE 1.2
Number of Items

	MC	CR	ER
ELA	57	2	1
Math	67	3	--

1.3 TECHNICAL REPORT CONTENT

This technical report summarizes the results of statistical and psychometric analyses performed on the Fall 2007, the Spring 2008, and the Summer 2008 operational data for the HSAP mathematics and ELA tests. All statistics are based on students in the regular schools only; students in adult education and district-approved homeschools are excluded. For Fall 2007 and Summer 2008, the data summary in all chapters of this technical report includes all students who attempted the HSAP tests. For Spring 2008, 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 2007 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 was caused by is due to the fact that some students and test administrators did not grid the grade field.

Table 2.1
Fall 2007 Summary of Student Demographics
in the HSAP Sample (All Attempts)

Demographic Category	Mathematics		ELA	
	N	%	N	%
All Students	13501	100.0	8081	100.0
Gender				
Female	6275	46.5	3224	39.9
Male	6628	49.1	4627	57.3
Invalid	598	4.4	230	2.8
Ethnicity				
African American	7787	57.7	4177	51.7
Asian/Pacific Islander	141	1.0	159	2.0
Hispanic	657	4.9	587	7.3
American Indian	15	0.1	18	0.2
White	4090	30.3	2732	33.8
Other	224	1.7	201	2.5
Invalid	587	4.3	207	2.6
Grade				
9	723	5.4	608	7.5
10	1849	13.7	1253	15.5
11	3651	27.0	2630	32.5
12	1501	11.1	1015	12.6
Invalid	5777	42.8	2575	31.9
ESL*				
No	4837	35.8	3452	42.7
Yes	466	3.5	448	5.5
Unknown	8198	60.7	4181	51.7
Lunch Program				
No free/reduced lunch	3123	23.1	2463	30.5
Free lunch	3176	23.5	2154	26.7
Reduced lunch	442	3.3	272	3.4
Unknown	6760	50.1	3192	39.5
IEP**				
No	11245	83.3	6462	80.0
Yes	2251	16.7	1612	19.9
Unknown	5	0.0	7	0.1
Migrant				
No	6509	48.2	4782	59.2
Yes	7	0.1	10	0.1
Unknown	6985	51.7	3289	40.7
Attempt				
1st	3054	22.6	2972	36.8
2nd	7579	56.1	3786	46.9
3rd	1358	10.1	571	7.1
4th or more	1150	8.5	513	6.3

* English as a second language

** individualized education program

Table 2.2
Spring 2008 Summary of Student Demographics
in the HSAP Sample (All Attempts)

Demographics	Mathematics		ELA	
	N	%	N	%
All Students	59229	100.0	56566	100.0
Gender				
Female	29514	49.8	27880	49.3
Male	29412	49.7	28449	50.3
Invalid	303	0.5	237	0.4
Ethnicity				
African American	24889	42.0	22968	40.6
Asian/Pacific Islander	717	1.2	740	1.3
Hispanic	2296	3.9	2285	4.0
American Indian	116	0.2	111	0.2
White	30113	50.8	29454	52.1
Other	793	1.3	769	1.4
Invalid	305	0.5	239	0.4
Grade				
9	6694	11.3	6591	11.7
10	47409	80.0	46814	82.8
11	3417	5.8	2182	3.9
12	1417	2.4	758	1.3
Invalid	292	0.5	221	0.4
ESL*				
No	55929	94.4	53416	94.4
Yes	1634	2.8	1697	3.0
Unknown	1666	2.8	1453	2.6
Lunch Program				
No free/reduced lunch	30647	51.7	29677	52.5
Free lunch	23730	40.1	22356	39.5
Reduced lunch	4250	7.2	4055	7.2
Unknown	602	1.0	478	0.8
IEP**				
No	50581	85.4	48696	86.1
Yes	8617	14.5	7837	13.9
Unknown	31	0.1	33	0.1
Migrant				
No	6479	10.9	6136	10.8
Yes	25	0.0	26	0.0
Unknown	52725	89.0	50404	89.1
Attempt				
1st	51661	87.2	51653	91.3
2nd	2348	4.0	1582	2.8
3rd	2807	4.7	1551	2.7
4th or more	1145	1.9	561	1.0

* English as a second language

** individualized education program

Table 2.3
Summer 2008 Summary of Student Demographics
in the HSAP Sample (All Attempts)

Demographics	Mathematics		ELA	
	N	%	N	%
All Students	237	100	96	100
Gender				
Female	108	45.6	37	38.5
Male	101	42.6	52	54.2
Invalid	28	11.8	7	7.3
Ethnicity				
African American	166	70.0	68	70.8
Asian/Pacific Islander	4	1.7	2	2.1
Hispanic	4	1.7	5	5.2
American Indian	0	0.0	0	0.0
White	30	12.7	12	12.5
Other	6	2.5	2	2.1
Invalid	27	11.4	7	7.3
Grade				
9	0	0.0	1	1.0
10	0	0.0	4	4.2
11	2	0.8	1	1.0
12	195	82.3	83	86.5
Invalid	40	16.9	7	7.3
ESL*				
No	114	48.1	48	50.0
Yes	3	1.3	3	3.1
Unknown	120	50.6	45	46.9
Lunch Program				
No free/reduced lunch	67	28.3	27	28.1
Free lunch	62	26.2	37	38.5
Reduced lunch	9	3.8	2	2.1
Unknown	99	41.8	30	31.3
IEP**				
No	86	36.3	30	31.3
Yes	151	63.7	66	68.8
Unknown	0	0.0	0	0.0
Migrant				
No	152	64.1	68	70.8
Yes	0	0.0	0	0.0
Unknown	85	35.9	28	29.2
Attempt				
1st	2	0.8	6	6.3
2nd	7	3.0	3	3.1
3rd	22	9.3	8	8.3
4th or more	203	85.7	76	79.2

* 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 2007a and 2008a). 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 2007 HSAP Administration (All Attempts)

Accommodation	Mathematics		ELA	
	Regular Form (N=13,474)	Customized Form (N=27)	Regular Form (N=8,054)	Customized Form (N=27)
Setting	16.8	88.9	19.6	96.3
Presentation	--	--	--	92.6
Timing	1.4	77.8	1.7	85.2
Schedule	0.4	37.0	0.4	40.7
Response options	0.8	70.4	1.0	85.2
Loose-leaf	0.2	--	0.1	--
Large-print	0.2	--	0.1	--
Spelling dictionary	--	--	1.4	40.7
Audiocassette	5.6	3.7	--	--
Oral script	5.9	7.4	--	--
Signed administration	--	88.9	--	--
Braille	--	11.1	--	7.4
Other	0.1	3.7	0.5	--

TABLE 2.5
Percentages of Students with Accommodations in the
Spring 2008 HSAP Administration (All Attempts)

Accommodation	Mathematics		ELA	
	Regular Form (N=59,178)	Customized Form (N=51)	Regular Form (N=56,515)	Customized Form (N=51)
Setting	7.9	76.5	8.1	88.2
Presentation	--	51.0	0.8	72.5
Timing (IEP)	0.5	51.0	0.7	58.8
Schedule (IEP)	0.1	45.1	0.2	56.9
Response options	0.1	11.8	0.5	64.7
Loose-leaf	0.0	--	0.0	2.0
Large-print	0.0	--	0.0	--
Spelling dictionary	--	--	0.5	54.9
Supplemental Materials	0.3	43.1	0.2	60.8
Audiocassette	2.7	--	--	--
Oral script	3.0	3.9	--	--
Signed administration	--	86.3	--	--
Braille	--	13.7	--	9.8
Bilingual Dictionary	0.8	--	1.2	--
Directions Translation	0.1	--	0.1	2.0
Individual/Small Group	0.7	7.8	0.8	2.0
Timing (ESL)	0.1	--	0.1	--
Schedule (ESL)	0.0	--	0.1	--
Other	0.1	11.8	0.1	11.8

TABLE 2.6
Percentages of Students with Accommodations in the
Summer 2008 Accommodations (All Attempts)

Accommodation	Mathematics		ELA	
	Regular Form (N=236)	Customized Form (N=1)	Regular Form (N=96)	Customized Form (N=0)
Setting	19.5	--	34.4	--
Presentation	--	--	6.3	--
Timing (IEP)	2.1	--	2.1	--
Schedule (IEP)	1.3	--	2.1	--
Response options	--	--	7.3	--
Loose-leaf	--	--	--	--
Large-print	--	--	--	--
Spelling dictionary	--	--	5.2	--
Supplemental Materials	5.1	--	--	--
Audiocassette	9.3	--	--	--
Oral script	18.6	--	--	--
Signed administration	--	100.0	--	--
Braille	--	--	--	--
Bilingual Dictionary	0.4	--	1.0	--
Directions Translation	--	--	--	--
Individual/Small Group	3.0	--	1.0	--
Timing (ESL)	--	--	--	--
Schedule (ESL)	--	--	--	--
Other	0.4	--	--	--

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 2007–08 HSAP Administrations (All Attempts)

Modification	Regular Form	Customized Form
Fall 2007		
	(N=8,054)	(N=27)
Alternative scoring	5.3	85.2
Extended writing option	1.2	70.4
Audiocassette	7.4	3.7
Oral administration	7.0	3.7
Signed administration	--	88.9
Spring 2008		
	(N=56,515)	(N=51)
Alternative scoring	2.0	70.6
Extended writing option	0.5	58.8
Audiocassette	2.9	--
Oral administration	2.5	2.0
Signed administration	--	84.3
Summer 2008		
	(N=96)	(N=0)
Alternative scoring	28.1	--
Extended writing option	11.5	--
Audiocassette	15.6	--
Oral administration	16.7	--
Signed administration	--	--

2.3 STUDENT QUESTIONNAIRES

After the test administration, students were instructed to answer the questions on the HSAP student questionnaire. Table 2.8 contains the number of questions on the student questionnaire. The questionnaire topics encompassed test difficulty, classroom activities, and (for mathematics only) calculator use.

TABLE 2.8
Number of Questions in the Student Questionnaire

	ELA	Math	Customized Math
Fall 2007	11	17	12
Spring 2008	11	17	12
Summer 2008	11	17	12

Chapter 3

TEST ADMINISTRATION

3.1 TEST ADMINISTRATION WINDOW

The HSAP ELA operational tests for Fall 2007 and Spring 2008 were conducted in two sessions over two days. The mathematics tests were conducted in one day. For the Summer 2008 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 2008 administration.

TABLE 3.1
2007–08 HSAP Test Administration Schedule

Fall 2007		Spring and Summer 2008	
Date	Test	Date	Test
October 23	ELA (day 1)	April 22	ELA (day 1)
October 24	ELA (day 2)	April 23	ELA (day 2)
October 25	Mathematics	April 24	Mathematics
October 26–November 2	Makeup tests window	April 25–May 2	Makeup tests window
		July 22–24	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 2007, Spring 2008, and Summer 2008 (All Attempts)

Time Taken	Fall 2007			Spring 2008			Summer 2008		
	Math % (N=13,501)	ELA % (N=8,081)		Math % (N=59,229)	ELA % (N=56,566)		Math % (N=237)	ELA % (N=96)	
		Session 1	Session 2		Session 1	Session 2		Session 1	Session 2
15 min	0.4	1.1	1.1	0.3	0.8	0.4	0.4	--	1.0
30 min	1.4	5.0	6.9	1.1	6.8	2.9	0.4	2.1	--
45 min	4.2	11.3	17.7	6.0	18.7	15.1	--	5.2	1.0
1 hr	10.8	18.1	20.6	18.0	24.0	25.8	3.0	18.8	7.3
1 hr 15 min	14.0	15.8	15.4	20.6	18.1	20.8	5.9	13.5	8.3
1 hr 30 min	14.7	14.1	10.5	17.7	11.9	13.3	8.9	13.5	11.5
1 hr 45 min	13.1	10.2	7.6	12.5	7.2	7.9	11.8	10.4	18.8
2 hr	11.6	7.0	5.0	8.6	4.3	4.9	18.1	9.4	13.5
2 hr 15 min	8.0	4.5	3.4	4.9	2.3	2.6	8.4	10.4	9.4
2 hr 30 min	5.3	3.0	2.1	2.9	1.2	1.3	8.9	5.2	4.2
2 hr 45 min	3.5	1.7	0.9	1.7	0.7	0.7	7.2	5.2	6.3
3 hr +	7.3	2.6	2.3	2.9	1.3	1.3	23.2	2.1	15.6
Invalid	5.7	5.6	6.3	2.7	2.6	2.9	3.8	4.2	3.1

3.3 ADMINISTRATION MANUALS

Working with SCDE staff, Pearson staff drafted the administration manuals for the test. SCDE staff reviewed and revised the manuals, and Pearson finalized and printed them. Two types of manuals were produced for the HSAP tests: the *HSAP Test Administration Manual (TAM)* and the *HSAP District Test Coordinator's Supplement* (SCDE 2007b and 2008b). The supplement included only the information that the DTCs needed for the administration of the HSAP tests. The *TAM* contained the information that the school test coordinators (STCs), TAs, and monitors needed to administer the tests to students in their schools. In addition, the Summer 2008 supplement (SCDE 2008c) was produced to provide specific information for the Summer 2008 administration and was used in conjunction with the information provided in the Spring 2008 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.

- Loose-leaf test booklets, which were printed, single sided, in three-ring binders, allowed individuals to remove the pages so that they could write or type answers to the constructed-response and extended-response items.
- Large-print booklets could be used for students who have difficulty reading text in a standard-size font. The large-print version was printed in a 9 x 12-inch spiral-bound booklet in an 18-point sans serif font.

- Braille booklets were produced for students who typically read classroom materials in braille. The braille versions were spiral bound on 11½ x 11-inch interpoint braille pages.
- A regular-print Form C test booklet was provided in test packets for students or the TAs to use with other customized formats such as the oral script/audiotape; braille, large-print, and loose-leaf versions; and sign language videotapes. These booklets were saddle-stitched and printed in a 12-point font just as the regular, noncustomized test booklets.
- Oral administration scripts and audiotapes were provided for students whose 504 and IEP plans were written to require oral administration of tests. Scripts provided the directions to the TAs regarding the appropriate way to read test questions, passages, and some answer choices to the students. Audiotapes were used for students testing individually or in small-group settings.
- Sign language videotapes were also produced and included the signed test directions, test questions, and some answer choices. The videotapes were produced in three languages: American Sign Language (ASL), Pidgin Signed English (PSE), and Signed Exact English (SEE).

3.5 PRETEST WORKSHOPS AND TRAINING

Pretest workshops were held in Columbia, South Carolina, to train the DTCs and some STCs. The DTCs were invited and could bring up to three additional representatives to the workshop. SCDE and Pearson staff trained the district staff in attendance.

Pearson 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.

The *DTC Supplement* and *TAMs* were mailed to the DTCs two weeks before the workshops and were also handed out to the DTCs during the workshop. The DTCs in attendance also received printed copies of the PowerPoint presentations used during the workshop. In addition, the PowerPoint presentations were posted to the SCDE Web site.

3.6 MATERIALS SHIPPING AND RETURN

Test materials were shipped to the district offices by Pearson and were scheduled to arrive at least three weeks before testing. Each school's shipment was boxed individually and labeled with the number of boxes shipped for that school. The Pearson shipment to each district office also included a 10 percent overage of all test materials—with the exception of customized formats, which were sent only in the quantities ordered. The 10 percent overage was in addition to the 5 percent overage included in school shipments. Overage materials for the districts were to be used by the DTCs to fulfill any additional materials requests from the STCs.

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

With the Pearson shipment of overage materials, the DTCs were sent “district coordinator kits,” which included step-by-step directions on how to return scorable and nonscorable materials. These directions listed toll-free phone numbers to call to schedule pickups of returned materials. The DTCs were given specific dates in the manuals for returning materials to Pearson. 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 to be encouraged to walk around the room during testing to check that students were marking their answers in the correct sections of the answer documents. It was permissible to alert students if they were marking their answers in the wrong sections of the answer documents. However, it was not permissible to stop and read test items or students’ responses in their test booklets.

Following the test administration and the return of materials, Pearson 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 Pearson attempted to contact the DTCs by telephone. Subsequently, the districts either located and returned the materials or sent explanations as to why materials were not found. A toll-free telephone number was provided to answer the DTCs’ questions regarding the missing materials; in addition, follow-up procedures were employed until all materials were accounted for.

Secure Materials

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

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

Chapter 4

SCORING

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 of a 0 to 3 scale. Condition codes of B (“blank”) and UR (“unreadable” or “illegible”) were used for nonscorable responses. For the purpose of calculating the total score, the condition codes were recoded as 0.

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

For the Graduation Express students, all 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.

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 2007, Spring and Summer 2008	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 2007					
Percentage	27%	17%	8%	8%	40%
Multiple-choice points	20	16	8	8	8
Constructed-response points	6	--	--	--	--
Extended-response points	--	--	--	--	30
Spring 2008					
Percentage	27%	17%	8%	8%	40%
Multiple-choice points	20	16	8	8	8
Constructed-response points	6	--	--	--	--
Extended-response points	--	--	--	--	30
Summer 2008					
Percentage	27%	17%	8%	8%	40%
Multiple-choice points	20	16	8	8	8
Constructed-response points	6	--	--	--	--
Extended-response points	--	--	--	--	30

4.3 SCORING PROCESS

Pearson scored all items: multiple-choice items were scored by Pearson’s electronic scanning system; constructed-response (CR) and extended-response (ER) items were scored by trained personnel using the ePEN (Electronic Performance Evaluation Network) system. There were two scoring sites for the HSAP in ELA: the Fall 2007 and Summer 2008 tests were scored in Iowa City, Iowa; the Spring 2008 test in Virginia Beach, VA. There was one scoring site for the HSAP in mathematics: all tests were scored in Lansing, Michigan.

Prior to actual scoring of the constructed-response and extended-response items, range-finding meetings were held in Columbia, South Carolina: September 11-13, 2007, for the Fall 2007 administration; January 15-17, for the Spring 2008 administration; and February 26-28, for the Summer 2008 administration. The purposes of the range-finding meetings were twofold: to identify sets of papers that were representative of the various performance levels defined by the rubric and to arrive at consensus scores on large sets of papers for use in training raters. Three range-finding committees—one each for reading, writing, and mathematics—were convened. The committees were composed of educators from South Carolina and were selected by the SCDE. Each committee reviewed several items. That is, each committee reviewed multiple papers (students’ responses written to a specific item) for multiple items.

SCDE staff were on-site during the first week of rater training (scorers received on-line training via the ePEN system) and live scoring and monitored the scoring process until scoring was complete. SCDE staff were on-site during the first week of the spring 2008 and summer 2008 training sessions. Throughout the scoring process, Pearson staff posted the performance of each reader (reader-reliability statistics) once a day on Pearson’s SchoolHouse Web site for SCDE staff to review.

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

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

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

4.4 READER RELIABILITY

In the scoring of constructed-response and extended-response items, 15 percent of the papers for CR items and 100 percent of the papers for ER items were independently scored by two readers. The percentages of reader consistency on the papers that were double-scored are reported in 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 2007			
CR1	2,027	84.3	99.8
CR2	2,038	94.2	99.3
CR3	2,024	93.4	99.7
Spring 2008			
CR1	7,762	89.2	99.5
CR2	7,711	88.4	99.5
CR3	7,736	91.2	99.6
Summer 2008			
CR1	39	97.4	97.4
CR2	37	97.3	97.3
CR3	36	94.4	100.0
ELA			
Fall 2007			
CR1	1,221	70.8	99.0
CR2	1,222	74.1	99.3
ER content and development	8,081	65.4	97.6
ER organization	8,081	62.7	96.9
ER voice	8,081	67.2	98.6
ER convention	8,081	57.3	93.7
Spring 2008			
CR1	7,699	66.3	97.2
CR2	7,730	69.5	98.6
ER content and development	51,527	73.3	98.7
ER organization	51,527	70.1	98.0
ER voice	51,527	76.8	99.3
ER convention	51,527	70.9	96.7
Summer 2008			
CR1	14	92.9	100.0
CR2	12	100.0	100.0
ER content and development	96	88.5	100.0
ER organization	96	82.3	100.0
ER voice	96	80.2	100.0
ER convention	96	74.0	99.0

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. After holding discussions, SCDE and Pearson staff 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 American Institute of Research (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 for mathematics, a multiple-choice (MC) item in both sessions 1 and 2 for ELA. The percentages in the adjacent column include students who omitted the last item on the test but answered the second-to-last item—CR item 2 for mathematics, item 19 in fall, item 14 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.

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

Subject	Last Item	Second-to-Last Item
Fall 2007		
Mathematics	91.8 (CR)	91.4 (CR)
ELA Session 1	98.9 (MC)	98.9 (MC)
ELA Session 2	98.3 (MC)	98.5 (MC)
Spring 2008		
Mathematics	94.0 (CR)	97.8 (CR)
ELA Session 1	99.2 (MC)	99.4 (MC)
ELA Session 2	99.4 (MC)	99.7 (MC)
Summer 2008		
Mathematics	96.6 (CR)	99.2 (CR)
ELA Session 1	100 (MC)	100 (MC)
ELA Session 2	100 (MC)	100 (MC)

5.2 CLASSICAL ITEM STATISTICS

Table 5.2 provides a summary of item p -values and item discriminations by item types and content areas for the mathematics operational items and a summary of item p -values and item discriminations by item types and content areas for the ELA operational and embedded field-test items. For 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 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
				Fall 2007	Spring 2008	Summer 2008			
Mathematics									
Multiple-choice	62	0.51	0.34	62	0.70	0.43	62	0.49	0.23
Constructed-response	3	0.38	0.57	3	0.63	0.57	3	0.34	0.35
Number and Operations	16	0.58	0.70	16	0.78	0.82	16	0.56	0.51
Algebra	19	0.54	0.72	19	0.70	0.81	19	0.51	0.53
Measurement and Geometry	19	0.42	0.73	19	0.61	0.79	19	0.39	0.48
Data Analysis and Probability	8	0.55	0.60	8	0.73	0.70	8	0.49	0.51
ELA									
Multiple-choice	60	0.55	0.37	60	0.71	0.37	60	0.51	0.24
Constructed-response	2	0.41	0.53	2	0.42	0.54	2	0.37	0.16
Extended-response	1	0.74	0.61	1	0.90	0.59	1	0.81	0.33
Reading Process and Comprehension	22	0.56	0.79	22	0.70	0.81	22	0.50	0.51
Analysis of Texts	16	0.52	0.74	16	0.69	0.75	16	0.49	0.52
Word Study and Analysis	8	0.58	0.72	8	0.76	0.64	8	0.55	0.50
Research	8	0.69	0.68	8	0.84	0.65	8	0.74	0.48
Writing	9	0.51	0.70	9	0.63	0.69	9	0.46	0.44

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 in the test booklet. 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:

Adequate: if $\theta \geq \theta_c + 1 \text{ SE}$

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

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

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 2007				
Number and Operations	0.044	0.568	-0.524	0.612
Algebra	-0.223	0.480	-0.703	0.257
Measurement and Geometry	-0.059	0.476	-0.535	0.417
Data Analysis and Probability	-0.115	0.733	-0.848	0.618
Spring 2008				
Number and Operations	0.009	0.558	-0.549	0.567
Algebra	0.010	0.498	-0.488	0.508
Measurement and Geometry	-0.056	0.483	-0.539	0.427
Data Analysis and Probability	-0.104	0.740	-0.844	0.636
Summer 2008				
Number and Operations	-0.008	0.548	-0.556	0.540
Algebra	-0.213	0.489	-0.702	0.276
Measurement and Geometry	-0.055	0.483	-0.538	0.428
Data Analysis and Probability	-0.049	0.742	-0.791	0.693
ELA				
Fall 2007				
Reading Process and Comprehension	0.019	0.442	-0.423	-0.461
Analysis of Texts	0.202	0.550	-0.348	0.752
Word Study and Analysis	0.102	0.753	-0.651	0.855
Research	0.109	0.721	-0.612	0.830
Writing	0.096	0.418	-0.322	0.514
Spring 2008				
Reading Process and Comprehension	0.141	0.428	-0.287	0.569
Analysis of Texts	0.186	0.523	-0.337	0.709
Word Study and Analysis	0.621	0.849	-0.228	1.470
Research	0.046	0.739	-0.693	0.785
Writing	0.071	0.399	-0.328	0.470
Summer 2008				
Reading Process and Comprehension	0.037	0.424	-0.387	0.461
Analysis of Texts	0.240	0.521	-0.281	0.761
Word Study and Analysis	0.193	0.765	-0.572	0.958
Research	0.550	0.751	-0.201	1.301
Writing	0.144	0.400	-0.256	0.544

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 2007	Spring 2008	Summer 2008
Percent Passed	35.0	81.0	8.0

TABLE 6.6
Fall 2007 HSAP Mathematics Operational Test: Percentage of Students
in Achievement Levels Overall and by Subgroups (All Attempts)

Subgroup	Achievement Levels				At or Above	At or Above	N
	Level 1	Level 2	Level 3	Level 4	Level 2	Level 3	
Overall	55.3	34.0	6.6	4.1	44.7	10.6	13,501
Gender							
Female	54.4	35.9	6.3	3.5	45.6	9.8	3,412
Male	58.0	32.5	5.7	3.8	42.0	9.5	3,846
Invalid	35.6	31.8	19.1	13.5	64.4	32.6	213
Ethnicity							
African American	64.5	32.6	2.4	0.5	35.5	2.9	5,022
Asian/Pacific Islander	36.2	29.1	17.7	17.0	63.8	34.8	51
Hispanic	57.7	34.2	5.9	2.1	42.3	8.1	379
American Indian	53.3	26.7	6.7	13.3	46.7	20.0	8
White	41.7	37.2	12.2	9.0	58.3	21.2	1,704
Other	41.1	34.4	13.8	10.7	58.9	24.6	92
Unknown	36.6	31.7	18.4	13.3	63.4	31.7	215
ESL*							
Yes	60.1	30.7	7.9	1.3	39.9	9.2	280
No	50.0	34.1	9.5	6.4	50.0	15.9	2,420
Unknown	58.2	34.2	4.8	2.9	41.8	7.6	4,771
Lunch Program							
No free/reduced lunch	40.1	33.4	14.8	11.7	59.9	26.5	1,251
Free lunch	61.9	32.7	4.1	1.4	38.1	5.4	1,965
Reduced lunch	50.7	37.6	8.4	3.4	49.3	11.8	224
Unknown	59.6	34.7	3.8	1.8	40.4	5.7	4,031
IEP**							
Yes	50.7	36.9	7.6	4.8	49.3	12.4	5,696
No	78.8	19.6	1.4	0.2	21.2	1.6	1,773
Unknown	40.0	60.0	--	--	60.0	0.0	2
Migrant							
Yes	28.6	71.4	--	--	71.4	0.0	2
No	51.6	32.5	9.5	6.4	48.4	15.9	3,359
Unknown	58.8	35.4	3.9	1.9	41.2	5.8	4,110

* English as a second language

** individualized education program

TABLE 6.7
Spring 2008 HSAP Mathematics Operational Test: Percentage of Students
in Achievement Levels Overall and by Subgroups (First Attempt)

Subgroup	Achievement Levels				At or Above	At or Above	N
	Level 1	Level 2	Level 3	Level 4	Level 2	Level 3	
Overall	15.5	28.5	32.7	23.3	84.5	56.0	51,661
Gender							
Female	13.4	29.8	33.4	23.4	86.6	56.8	3,484
Male	17.4	27.2	32.1	23.3	82.6	55.4	4,428
Invalid	33.3	29.7	26.0	11.0	66.7	37.0	73
Ethnicity							
African American	25.8	38.0	27.2	9.1	74.2	36.3	5,127
Asian/Pacific Islander	4.6	15.1	28.9	51.4	95.4	80.4	30
Hispanic	21.3	31.1	31.5	16.1	78.7	47.6	405
American Indian	17.0	30.0	31.0	22.0	83.0	53.0	17
White	7.9	21.9	36.9	33.3	92.1	70.2	2,235
Other	15.1	30.6	31.6	22.8	84.9	54.4	103
Unknown	30.2	30.7	26.2	12.9	69.8	39.1	68
ESL*							
Yes	25.7	31.3	29.0	14.0	74.3	43.1	331
No	14.8	28.3	33.0	23.9	85.2	56.9	7,263
Unknown	32.2	33.0	25.2	9.6	67.8	34.8	391
Lunch Program							
No free/reduced lunch	8.2	22.6	35.7	33.5	91.8	69.2	2,310
Free lunch	25.5	36.4	28.1	10.0	74.5	38.0	4,974
Reduced lunch	15.1	31.8	35.1	18.0	84.9	53.1	561
Unknown	33.6	31.9	24.5	10.1	66.4	34.5	140
IEP**							
Yes	10.0	28.5	35.4	26.0	90.0	61.4	4,585
No	57.4	28.4	11.9	2.3	42.6	14.2	3,388
Unknown	60.0	25.0	10.0	5.0	40.0	15.0	12
Migrant							
Yes	33.3	23.8	28.6	14.3	66.7	42.9	7
No	30.1	32.5	26.2	11.2	69.9	37.4	1,586
Unknown	13.8	28.1	33.4	24.7	86.2	58.1	6,392

* English as a second language
** individualized education program

TABLE 6.8
Summer 2008 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	65.4	32.9	1.3	0.4	34.6	1.7	237
Gender							
Female	67.6	32.4	--	--	32.4	0.0	73
Male	67.3	29.7	2.0	1.0	32.7	3.0	68
Invalid	50.0	46.4	3.6	--	50.0	3.6	14
Ethnicity							
African American	70.5	28.3	1.2	--	29.5	1.2	117
Asian/Pacific Islander	100.0	--	--	--	0.0	0.0	4
Hispanic	100.0	--	--	--	0.0	0.0	4
American Indian	--	--	--	--	0.0	0.0	--
White	46.7	50.0	3.3	--	53.3	3.3	14
Other	50.0	50.0	--	--	50.0	0.0	3
Unknown	48.1	48.1	3.7	--	51.9	3.7	13
ESL*							
Yes	100.0	--	--	--	0.0	0.0	3
No	62.3	37.7	--	--	37.7	0.0	71
Unknown	67.5	29.2	2.5	0.8	32.5	3.3	81
Lunch Program							
No free/reduced lunch	70.1	28.4	1.5	--	29.9	1.5	47
Free lunch	64.5	33.9	1.6	--	35.5	1.6	40
Reduced lunch	66.7	33.3	--	--	33.3	0.0	6
Unknown	62.6	35.4	2.0	--	37.4	2.0	62
IEP**							
Yes	59.3	40.7	--	--	40.7	0.0	51
No	68.9	28.5	2.0	0.7	31.1	2.6	104
Unknown	--	--	--	--	0.0	0.0	--
Migrant							
Yes	--	--	--	--	0.0	0.0	--
No	67.1	32.9	--	--	32.9	0.0	102
Unknown	62.4	32.9	3.5	1.2	37.6	4.7	53

* English as a second language
** individualized education program

TABLE 6.9
Fall 2007 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	54.4	26.2	10.8	8.6	45.6	19.5	8,081
Gender							
Female	48.8	26.6	13.7	11.0	51.2	24.7	1,573
Male	59.2	26.0	8.3	6.6	40.8	14.9	2,738
Invalid	36.1	25.2	21.3	17.4	63.9	38.7	83
Ethnicity							
African American	68.4	25.1	4.8	1.7	31.6	6.5	2,855
Asian/Pacific Islander	49.1	32.1	11.3	7.5	50.9	18.9	78
Hispanic	65.8	22.0	8.2	4.1	34.2	12.3	386
American Indian	50.0	33.3	5.6	11.1	50.0	16.7	9
White	33.5	28.7	18.9	19.0	66.5	37.8	914
Other	37.8	20.9	22.9	18.4	62.2	41.3	76
Unknown	36.7	25.6	21.3	16.4	63.3	37.7	76
ESL*							
Yes	66.5	25.0	6.5	2.0	33.5	8.5	298
No	42.8	28.0	15.6	13.5	57.2	29.1	1,479
Unknown	62.6	24.8	7.3	5.3	37.4	12.7	2,617
Lunch Program							
No free/reduced lunch	29.6	27.5	22.0	20.9	70.4	42.9	729
Free lunch	62.8	27.7	6.8	2.8	37.2	9.6	1,352
Reduced lunch	53.7	26.1	13.6	6.6	46.3	20.2	146
Unknown	67.9	24.1	4.7	3.3	32.1	8.0	2,167
IEP**							
Yes	48.2	28.1	13.0	10.7	51.8	23.7	3,116
No	79.0	18.6	2.1	0.3	21.0	2.4	1,273
Unknown	71.4	14.3	14.3	--	28.6	14.3	5
Migrant							
Yes	90.0	10.0	--	--	10.0	0.0	9
No	45.2	27.7	15.0	12.2	54.8	27.2	2,160
Unknown	67.6	24.0	4.8	3.5	32.4	8.3	2,225

* English as a second language

**individualized education program

TABLE 6.10
Spring 2008 HSAP ELA Operational Test: Percentage of Students
in Achievement Levels Overall and by Subgroups (First Attempt)

Subgroup	Achievement Levels				At or Above Level 2	At or Above Level 3	N
	Level 1	Level 2	Level 3	Level 4			
Overall	12.2	28.5	34.2	25.1	87.8	59.2	51,653
Gender							
Female	8.4	26.7	35.6	29.2	91.6	64.8	2,186
Male	16.0	30.4	32.7	20.9	84.0	53.6	4,080
Invalid	30.5	25.8	30.5	13.2	69.5	43.7	58
Ethnicity							
African American	19.4	40.4	30.5	9.7	80.6	40.2	3,871
Asian/Pacific Islander	8.3	18.6	33.3	39.8	91.7	73.1	54
Hispanic	22.2	29.8	31.7	16.3	77.8	48.0	423
American Indian	12.9	34.7	32.7	19.8	87.1	52.5	13
White	6.5	20.3	37.0	36.2	93.5	73.2	1,832
Other	11.4	26.8	33.7	28.1	88.6	61.8	78
Unknown	27.7	27.7	29.3	15.2	72.3	44.5	53
ESL*							
Yes	31.5	31.9	25.3	11.3	68.5	36.6	406
No	11.4	28.3	34.6	25.7	88.6	60.3	5,601
Unknown	26.8	34.1	27.1	12.1	73.2	39.2	317
Lunch Program							
No free/reduced lunch	6.0	20.6	36.7	36.7	94.0	73.4	1,688
Free lunch	21.0	39.1	29.9	10.0	79.0	39.8	4,104
Reduced lunch	11.3	32.7	38.5	17.5	88.7	56.0	423
Unknown	29.7	32.2	26.2	12.0	70.3	38.1	109
IEP**							
Yes	7.2	27.6	37.0	28.1	92.8	65.1	3,301
No	50.2	35.5	12.3	2.0	49.8	14.3	3,009
Unknown	58.3	25.0	16.7	--	41.7	16.7	14
Migrant							
Yes	38.1	28.6	14.3	19.0	61.9	33.3	8
No	25.5	33.9	27.2	13.3	74.5	40.5	1,350
Unknown	10.7	27.9	35.0	26.4	89.3	61.4	4,966

* English as a second language

** individualized education program

TABLE 6.11
Summer 2008 HSAP ELA 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	59.4	34.4	5.2	1.0	40.6	6.3	96
Gender							
Female	70.3	27.0	2.7	--	29.7	2.7	26
Male	48.1	42.3	7.7	1.9	51.9	9.6	25
Invalid	85.7	14.3	--	--	14.3	0.0	6
Ethnicity							
African American	66.2	32.4	1.5	--	33.8	1.5	45
Asian/Pacific Islander	50.0	50.0	--	--	50.0	0.0	1
Hispanic	60.0	40.0	--	--	40.0	0.0	3
American Indian	--	--	--	--	0.0	0.0	--
White	33.3	33.3	25.0	8.3	66.7	33.3	4
Other	100.0	--	--	--	0.0	0.0	2
Unknown	57.1	42.9	--	--	42.9	0.0	4
ESL*							
Yes	66.7	33.3	--	--	33.3	0.0	2
No	60.4	33.3	6.3	--	39.6	6.3	29
Unknown	57.8	37.8	2.2	2.2	42.2	4.4	26
Lunch Program							
No free/reduced lunch	55.6	29.6	11.1	3.7	44.4	14.8	15
Free lunch	67.6	29.7	2.7	--	32.4	2.7	25
Reduced lunch	50.0	50.0	--	--	50.0	0.0	1
Unknown	53.3	46.7	--	--	46.7	0.0	16
IEP**							
Yes	60.0	36.7	3.3	--	40.0	3.3	18
No	59.1	33.3	6.1	1.5	40.9	7.6	39
Unknown	--	--	--	--	0.0	0.0	--
Migrant							
Yes	--	--	--	--	0.0	0.0	--
No	63.2	30.9	5.9	--	36.8	5.9	43
Unknown	50.0	42.9	3.6	3.6	50.0	7.1	14

* English as a second language

** individualized education program

TABLE 6.12
Fall 2007 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	80.7%	18.4%	0.9%	7,471	21.8%	43.3%	34.9%	6,030
Females	84.1%	15.4%	0.5%	3,412	28.2%	42.8%	29.1%	2,863
Males	77.8%	21.0%	1.2%	3,846	17.0%	46.0%	37.1%	2,782
African Americans	83.2%	16.3%	0.5%	5,022	31.5%	50.0%	18.5%	2,765
Whites	73.7%	24.6%	1.7%	1,704	13.7%	37.6%	48.7%	2,386
Algebra								
All students	37.8%	58.9%	3.3%	7,471	1.6%	50.0%	48.4%	6,030
Females	32.8%	63.0%	4.2%	3,412	1.3%	48.9%	49.8%	2,863
Males	42.2%	55.4%	2.5%	3,846	1.9%	53.8%	44.3%	2,782
African Americans	36.2%	60.3%	3.5%	5,022	1.7%	59.1%	39.2%	2,765
Whites	42.3%	55.5%	2.2%	1,704	1.4%	44.6%	54.0%	2,386
Measurement and Geometry								
All students	63.4%	36.3%	0.3%	7,471	7.1%	59.1%	33.8%	6,030
Females	62.8%	37.0%	0.1%	3,412	7.6%	62.6%	29.8%	2,863
Males	63.9%	35.6%	0.5%	3,846	6.9%	58.7%	34.4%	2,782
African Americans	63.9%	35.9%	0.2%	5,022	9.9%	71.5%	18.6%	2,765
Whites	60.8%	38.5%	0.7%	1,704	4.7%	49.3%	46.0%	2,386
Data Analysis and Probability								
All students	25.4%	65.6%	9.0%	7,471	1.5%	46.5%	52.0%	6,030
Females	21.2%	67.8%	10.9%	3,412	0.9%	45.5%	53.5%	2,863
Males	29.3%	63.5%	7.2%	3,846	2.2%	48.9%	48.9%	2,782
African Americans	25.4%	65.1%	9.4%	5,022	1.8%	51.3%	46.9%	2,765
Whites	25.6%	66.4%	8.0%	1,704	1.1%	41.8%	57.1%	2,386

* total number of students in Level 1

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

TABLE 6.13
Spring 2008 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>	N1*	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	N2**
Number and Operations								
All students	85.4%	13.8%	0.8%	7,985	6.0%	23.8%	70.3%	43,676
Females	86.2%	13.5%	0.3%	3,484	7.1%	26.3%	66.5%	22,503
Males	84.8%	14.2%	1.1%	4,428	4.7%	21.0%	74.3%	21,027
African Americans	86.9%	12.5%	0.6%	5,127	11.0%	35.1%	53.9%	14,777
Whites	82.1%	16.7%	1.2%	2,235	3.1%	17.2%	79.6%	25,958
Algebra								
All students	57.8%	40.0%	2.2%	7,985	1.7%	25.4%	73.0%	43,676
Females	51.9%	45.4%	2.8%	3,484	1.5%	24.5%	74.0%	22,503
Males	62.5%	35.8%	1.7%	4,428	1.9%	26.3%	71.8%	21,027
African Americans	55.8%	42.1%	2.1%	5,127	2.2%	34.3%	63.6%	14,777
Whites	63.0%	35.0%	2.0%	2,235	1.4%	20.6%	78.0%	25,958
Measurement and Geometry								
All students	55.8%	43.9%	0.4%	7,985	2.8%	37.8%	59.4%	43,676
Females	56.4%	43.3%	0.2%	3,484	3.2%	39.5%	57.3%	22,503
Males	55.3%	44.3%	0.4%	4,428	2.3%	36.0%	61.7%	21,027
African Americans	57.1%	42.6%	0.3%	5,127	4.9%	54.4%	40.8%	14,777
Whites	52.3%	47.1%	0.5%	2,235	1.6%	28.6%	69.8%	25,958
Data Analysis and Probability								
All students	31.8%	60.3%	7.9%	7,985	0.5%	23.1%	76.3%	43,676
Females	26.8%	64.0%	9.2%	3,484	0.4%	22.5%	77.1%	22,503
Males	35.6%	57.3%	7.0%	4,428	0.6%	23.8%	75.6%	21,027
African Americans	31.0%	60.8%	8.3%	5,127	0.8%	31.0%	68.2%	14,777
Whites	34.2%	58.5%	7.3%	2,235	0.3%	18.2%	81.5%	25,958

* total number of students in Level 1

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

TABLE 6.14
Summer 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	73.5%	25.8%	0.6%	155	17.1%	57.3%	25.6%	82
Females	72.6%	26.0%	1.4%	73	17.1%	74.3%	8.6%	35
Males	75.0%	25.0%	--	68	9.1%	45.5%	45.5%	33
African Americans	72.6%	26.5%	0.9%	117	14.3%	61.2%	24.5%	49
Whites	71.4%	28.6%	--	14	6.3%	56.3%	37.5%	16
Algebra								
All students	29.7%	67.1%	3.2%	155	--	58.5%	41.5%	82
Females	17.8%	79.5%	2.7%	73	--	57.1%	42.9%	35
Males	38.2%	57.4%	4.4%	68	--	66.7%	33.3%	33
African Americans	30.8%	65.8%	3.4%	117	--	69.4%	30.6%	49
Whites	21.4%	78.6%	--	14	--	43.8%	56.3%	16
Measurement and Geometry								
All students	49.7%	50.3%	--	155	11.0%	79.3%	9.8%	82
Females	46.6%	53.4%	--	73	2.9%	94.3%	2.9%	35
Males	50.0%	50.0%	--	68	15.2%	66.7%	18.2%	33
African Americans	51.3%	48.7%	--	117	10.2%	79.6%	10.2%	49
Whites	21.4%	78.6%	--	14	6.3%	81.3%	12.5%	16
Data Analysis and Probability								
All students	34.2%	61.3%	4.5%	155	4.9%	51.2%	43.9%	82
Females	32.9%	63.0%	4.1%	73	5.7%	45.7%	48.6%	35
Males	33.8%	60.3%	5.9%	68	6.1%	54.5%	39.4%	33
African Americans	35.9%	59.8%	4.3%	117	6.1%	57.1%	36.7%	49
Whites	28.6%	64.3%	7.1%	14	--	43.8%	56.3%	16

* total number of students in Level 1

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

TABLE 6.15
Fall 2007 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	47.7%	46.4%	6.0%	4,394	1.2%	27.9%	70.8%	3,687
Females	41.6%	51.3%	7.1%	1,573	0.9%	23.6%	75.5%	1,651
Males	51.0%	43.6%	5.4%	2,738	1.6%	32.5%	65.9%	1,889
African Americans	47.0%	47.6%	5.4%	2,855	2.0%	39.3%	58.7%	1,322
Whites	50.0%	44.1%	5.9%	914	0.9%	21.8%	77.3%	1,818
Analysis of Texts								
All students	70.0%	26.9%	3.1%	4,394	9.5%	38.1%	52.4%	3,687
Females	70.4%	26.9%	2.7%	1,573	9.3%	35.9%	54.9%	1,651
Males	69.9%	26.9%	3.3%	2,738	10.4%	40.6%	49.0%	1,889
African Americans	71.0%	26.4%	2.6%	2,855	14.6%	51.0%	34.4%	1,322
Whites	64.7%	30.9%	4.5%	914	5.8%	30.0%	64.2%	1,818
Word Study and Analysis								
All students	52.0%	45.4%	2.6%	4,394	5.6%	51.5%	42.9%	3,687
Females	54.8%	43.6%	1.6%	1,573	7.0%	50.0%	43.1%	1,651
Males	50.5%	46.3%	3.2%	2,738	4.7%	53.1%	42.2%	1,889
African American	53.6%	44.3%	2.2%	2,855	8.8%	66.0%	25.3%	1,322
White	46.1%	49.3%	4.6%	914	3.1%	41.7%	55.2%	1,818
Writing								
All students	39.9%	55.8%	4.3%	4,394	4.4%	45.2%	50.4%	3,687
Females	40.6%	55.6%	3.8%	1,573	4.3%	44.3%	51.4%	1,651
Males	39.3%	56.0%	4.7%	2,738	4.4%	47.0%	48.5%	1,889
African Americans	40.5%	55.3%	4.3%	2,855	6.3%	58.9%	34.8%	1,322
Whites	39.1%	55.6%	5.4%	914	3.2%	36.4%	60.4%	1,818
Research								
All students	75.6%	22.1%	2.3%	4,394	6.5%	26.9%	66.7%	3,687
Females	71.3%	26.4%	2.4%	1,573	4.7%	23.6%	71.7%	1,651
Males	77.9%	19.7%	2.3%	2,738	8.2%	30.3%	61.5%	1,889
African Americans	75.3%	22.5%	2.2%	2,855	9.3%	36.5%	54.2%	1,322
Whites	72.4%	24.2%	3.4%	914	4.5%	20.2%	75.3%	1,818

* total number of students in Level 1

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

TABLE 6.16
Spring 2008 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	62.3%	35.0%	2.7%	6,324	1.2%	18.1%	80.7%	45,329
Females	60.0%	37.6%	2.5%	2,186	1.1%	16.3%	82.7%	23,738
Males	63.5%	33.7%	2.8%	4,080	1.3%	20.2%	78.4%	21,459
African Americans	61.2%	36.7%	2.1%	3,871	2.0%	28.3%	69.7%	16,103
Whites	63.5%	32.3%	4.3%	1,832	0.7%	11.8%	87.5%	26,313
Analysis of Texts								
All students	64.5%	29.0%	6.5%	6,324	3.3%	15.4%	81.3%	45,329
Females	60.8%	33.0%	6.2%	2,186	2.7%	13.9%	83.4%	23,738
Males	66.3%	26.9%	6.7%	4,080	4.0%	16.9%	79.0%	21,459
African Americans	64.8%	29.5%	5.7%	3,871	5.5%	22.8%	71.7%	16,103
Whites	63.0%	28.4%	8.6%	1,832	2.0%	10.8%	87.2%	26,313
Word Study and Analysis								
All students	63.7%	31.5%	4.8%	6,324	7.3%	41.5%	51.2%	45,329
Females	61.6%	33.4%	5.0%	2,186	7.3%	42.9%	49.8%	23,738
Males	64.9%	30.5%	4.7%	4,080	7.4%	39.9%	52.7%	21,459
African Americans	65.8%	30.1%	4.1%	3,871	11.5%	49.2%	39.3%	16,103
Whites	58.5%	35.3%	6.2%	1,832	4.8%	36.8%	58.4%	26,313
Writing								
All students	23.8%	67.9%	8.3%	6,324	1.5%	27.8%	70.7%	45,329
Females	24.4%	68.4%	7.2%	2,186	1.5%	27.8%	70.7%	23,738
Males	23.5%	67.7%	8.8%	4,080	1.5%	27.8%	70.7%	21,459
African Americans	23.6%	68.7%	7.6%	3,871	2.3%	38.5%	59.2%	16,103
Whites	24.0%	66.3%	9.8%	1,832	1.0%	21.1%	78.0%	26,313
Research								
All students	62.5%	31.1%	6.3%	6,324	2.0%	15.6%	82.5%	45,329
Females	57.1%	35.9%	7.0%	2,186	1.2%	12.9%	85.9%	23,738
Males	65.3%	28.6%	6.1%	4,080	2.8%	18.5%	78.7%	21,459
African Americans	60.7%	32.7%	6.6%	3,871	2.9%	22.8%	74.3%	16,103
Whites	62.5%	30.5%	7.0%	1,832	1.4%	11.0%	87.6%	26,313

* total number of students in Level 1

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

TABLE 6.17
Summer 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	35.1%	61.4%	3.5%	57	7.7%	46.2%	46.2%	39
Females	30.8%	65.4%	3.8%	26	9.1%	45.5%	45.5%	11
Males	40.0%	56.0%	4.0%	25	7.4%	44.4%	48.1%	27
African Americans	35.6%	60.0%	4.4%	45	8.7%	52.2%	39.1%	23
Whites	50.0%	50.0%	--	4	--	37.5%	62.5%	8
Analysis of Texts								
All students	66.7%	31.6%	1.8%	57	25.6%	43.6%	30.8%	39
Females	69.2%	26.9%	3.8%	26	27.3%	54.5%	18.2%	11
Males	60.0%	40.0%	--	25	25.9%	37.0%	37.0%	27
African Americans	64.4%	33.3%	2.2%	45	30.4%	47.8%	21.7%	23
Whites	100.0%	--	--	4	--	37.5%	62.5%	8
Word Study and Analysis								
All students	49.1%	49.1%	1.8%	57	10.3%	56.4%	33.3%	39
Females	53.8%	46.2%	--	26	9.1%	63.6%	27.3%	11
Males	48.0%	48.0%	4.0%	25	7.4%	55.6%	37.0%	27
African American	48.9%	48.9%	2.2%	45	13.0%	56.5%	30.4%	23
White	25.0%	75.0%	--	4	--	50.0%	50.0%	8
Writing								
All students	64.9%	33.3%	1.8%	57	23.1%	66.7%	10.3%	39
Females	61.5%	38.5%	--	26	18.2%	72.7%	9.1%	11
Males	68.0%	28.0%	4.0%	25	25.9%	63.0%	11.1%	27
African Americans	62.2%	35.6%	2.2%	45	34.8%	60.9%	4.3%	23
Whites	75.0%	25.0%	--	4	12.5%	50.0%	37.5%	8
Research								
All students	64.9%	29.8%	5.3%	57	15.4%	38.5%	46.2%	39
Females	73.1%	23.1%	3.8%	26	--	54.5%	45.5%	11
Males	64.0%	32.0%	4.0%	25	22.2%	33.3%	44.4%	27
African Americans	71.1%	24.4%	4.4%	45	21.7%	43.5%	34.8%	23
Whites	25.0%	75.0%	--	4	12.5%	25.0%	62.5%	8

* 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

Subgroup	Mathematics			ELA		
	N	Scale Score		N	Scale Score	
		Mean	Std. Dev.		Mean	Std. Dev.
Fall 2007						
All students	13,501	199.0	19.3	8,081	201.7	25.0
Females	6,275	199.4	17.9	3,224	205.9	25.4
Males	6,628	197.6	19.5	4,627	198.3	23.9
African Americans	7,787	194.1	13.9	4,177	193.5	18.5
Whites	4,090	206.0	22.9	2,732	214.1	27.8
Spring 2008						
All students	51,661	223.4	25.3	51,653	226.0	22.6
Females	25,987	224.3	24.2	25,924	229.4	21.5
Males	25,455	222.6	26.3	25,539	222.7	23.2
African Americans	19,904	212.3	21.7	19,974	216.4	20.2
Whites	28,193	231.3	24.5	28,145	233.3	21.3
Summer 2008						
All students	237	195.8	11.3	96	199.1	12.9
Females	108	195.3	9.8	37	196.9	11.0
Males	101	196.0	12.5	52	201.4	14.4
African Americans	166	194.6	10.7	68	196.4	10.2
Whites	30	201.4	13.2	12	213.3	19.4

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 2007	Spring 2008	Summer 2008
Mathematics			
Reliability	0.89	0.93	0.76
SEM	3.69	3.41	3.75
ELA			
Reliability	0.95	0.94	0.84
SEM	4.99	4.01	4.36

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 2007	6.35	5.52	5.98	7.43
Mathematics, Spring 2008	6.73	5.63	5.93	7.20
Mathematics, Summer 2008	5.59	5.51	5.95	7.22
ELA, Fall 2007	7.24	5.36	6.12	7.39
ELA, Spring 2008	6.52	5.24	5.98	7.23
ELA, Summer 2008	5.69	5.15	6.11	7.50

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 bivariate normal method (Huynh and Kim 2008). 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 2007		
Mathematics	85.0%	91.8%
ELA	89.9%	93.0%
Spring 2008		
Mathematics	92.2%	88.1%
ELA	94.3%	89.0%
Summer 2008		
Mathematics	79.0%	97.9%
ELA	81.8%	96.7%

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 Pearson selected for detecting DIF were the Mantel-Haenszel (MH) chi-square for dichotomous items (MC items) and Mantel's chi-square for polytomous items (CR and ER items). Pearson 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.

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 2007, Spring 2008, and Summer 2008 operational forms, Tables 8.1, 8.2 and 8.3 report the numbers of items in the various DIF categories for mathematics and ELA.

TABLE 8.1
Fall 2007 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	4	--
Multiple choice	White	Black	62	57	4	1
Constructed response	Male	Female	3	3	--	--
Constructed response	White	Black	3	3	--	--
ELA						
Multiple choice	Male	Female	60	59	1	--
Multiple choice	White	Black	60	57	3	--
Constructed response	Male	Female	2	--	1	1
Constructed response	White	Black	2	1	1	--
Extended response	Male	Female	1	1	--	--
Extended response	White	Black	1	1	--	--

TABLE 8.2
Spring 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	53	8	1
Multiple choice	White	Black	62	54	8	--
Constructed response	Male	Female	3	3	--	--
Constructed response	White	Black	3	3	--	--
ELA						
Multiple choice	Male	Female	60	58	2	--
Multiple choice	White	Black	60	58	2	--
Constructed response	Male	Female	2	1	1	--
Constructed response	White	Black	2	2	--	--
Extended response	Male	Female	1	1	--	--
Extended response	White	Black	1	1	--	--

TABLE 8.3
Summer 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	62	--	--
Multiple choice	White	Black	62	61	--	1
Constructed response	Male	Female	3	3	--	--
Constructed response	White	Black	3	3	--	--
ELA						
Multiple choice	Male	Female	60	60	--	--
Multiple choice	White	Black	60	60	--	--
Constructed response	Male	Female	2	2	--	--
Constructed response	White	Black	2	2	--	--
Extended response	Male	Female	1	1	--	--
Extended response	White	Black	1	1	--	--

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 2007 HSAP Correlations among Reporting Categories

Mathematics (N=13,501)						ELA (N=8,081)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
NO	1.00	0.60	0.63	0.47	0.58	RC	1.00	0.72	0.68	0.66	0.67
AL	--	1.00	0.62	0.52	0.60	AT	--	1.00	0.67	0.61	0.59
MG	--	--	1.00	0.52	0.60	WS	--	--	1.00	0.59	0.58
DP	--	--	--	1.00	0.51	WR	--	--	--	1.00	0.55
IR	--	--	--	--	1.00	RS	--	--	--	--	1.00

TABLE 8.5
Spring 2008 HSAP Correlations among Reporting Categories

Mathematics (N=51,661)						ELA (N=51,653)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
NO	1.00	0.75	0.73	0.66	0.68	RC	1.00	0.73	0.62	0.63	0.67
AL	--	1.00	0.74	0.62	0.66	AT	--	1.00	0.56	0.60	0.59
MG	--	--	1.00	0.60	0.64	WS	--	--	1.00	0.47	0.52
DP	--	--	--	1.00	0.60	WR	--	--	--	1.00	0.50
IR	--	--	--	--	1.00	RS	--	--	--	--	1.00

TABLE 8.6
Summer 2008 HSAP Correlations among Reporting Categories

Mathematics (N=237)						ELA (N=96)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
NO	1.00	0.43	0.34	0.37	0.32	RC	1.00	0.45	0.41	0.39	0.32
AL	--	1.00	0.36	0.43	0.30	AT	--	1.00	0.47	0.34	0.32
MG	--	--	1.00	0.34	0.41	WS	--	--	1.00	0.25	0.31
DP	--	--	--	1.00	0.32	WR	--	--	--	1.00	0.38
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, and Do-Hong Kim. 2008. "Assessing Reliability and Classification Consistency for Group-Based Index with Applications to School Accountability." Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.
- 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, Geoff 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. 2007a. *Fall 2007 HSAP Test Administration Manual*. Columbia: South Carolina Department of Education.

- . 2007b. *Fall 2007 HSAP District Test Coordinator's Supplement*. Columbia: South Carolina Department of Education.
- . 2008a. *Spring 2008 HSAP Test Administration Manual*. Columbia: South Carolina Department of Education.
- . 2008b. *Spring 2008 HSAP District Test Coordinator's Supplement*. Columbia: South Carolina Department of Education.
- . 2008c. *Summer 2008 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.