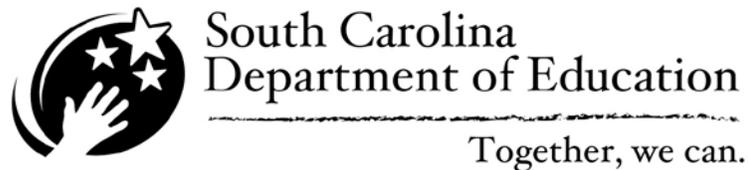


**South Carolina  
High School Assessment Program**

**English Language Arts and Mathematics  
2006–07 Operational Test Technical Report**



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.....	43
CHAPTER 7. RELIABILITY .....		44
7.1	Reliability of Raw Scores .....	44
7.2	Overall and Conditional SEM.....	44
7.3	Consistency of Achievement Levels.....	45
CHAPTER 8. VALIDITY .....		46
8.1	Item Distribution across Strands.....	46
8.2	Item Development.....	46
8.3	Differential Item Functioning (DIF).....	46
	Procedure .....	47
8.4	Correlations among Reporting Categories.....	49
REFERENCES .....		51

## LIST OF TABLES

2.1	Fall 2006 Summary of Student Demographics in the HSAP Sample (All Attempts).....	4
2.2	Spring 2007 Summary of Student Demographics in the HSAP Sample (All Attempts) .....	5
2.3	Summer 2007 Summary of Student Demographics in the HSAP Sample (All Attempts)...	6
2.4	Percentages of Students with Accommodations in the Fall 2006 HSAP Administration (All Attempts).....	8
2.5	Percentages of Students with Accommodations in the Spring 2007 HSAP Administration (All Attempts) .....	8
2.6	Percentages of Students with Accommodations in the Summer 2007 HSAP Administration (All Attempts) .....	9
2.7	Percentages of Students with Modifications in the 2006–07 HSAP Administrations (All Attempts) .....	10
3.1	2006–07 HSAP Test Administration Schedule.....	11
3.2	Percentage of Students by Test Duration: HSAP Fall 2006, Spring 2007, and Summer 2007 (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	Fall 2006 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	31
6.6	Spring 2007 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (First Attempt).....	32
6.7	Summer 2007 HSAP Mathematics Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	33

6.8	Fall 2006 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	34
6.9	Spring 2007 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (First Attempt) .....	35
6.10	Summer 2007 HSAP ELA Operational Test: Percentage of Students in Achievement Levels Overall and by Subgroups (All Attempts).....	36
6.11	Fall 2006 HSAP Mathematics Operational Test: Content-Area Information (All Attempts).....	37
6.12	Spring 2007 HSAP Mathematics Operational Test: Content-Area Information (First Attempt) .....	38
6.13	Summer 2007 HSAP Mathematics Operational Test: Content-Area Information (All Attempts).....	39
6.14	Fall 2006 HSAP ELA Operational Test: Content-Area Information (All Attempts).....	40
6.15	Spring 2007 HSAP ELA Operational Test: Content-Area Information (First Attempt) .....	41
6.16	Summer 2007 HSAP ELA Operational Test: Content-Area Information (All Attempts).....	42
6.17	HSAP Summary Statistics Overall and by Subgroups .....	43
7.1	Reliability Coefficients and SEM for HSAP Raw Scores.....	44
7.2	Classical and Conditional SEM for HSAP .....	45
7.3	Consistency Indexes for HSAP Achievement Levels .....	45
8.1	Fall 2006 Summary of DIF Classifications for HSAP Mathematics and ELA Items .....	48
8.2	Spring 2007 Summary of DIF Classifications for HSAP Mathematics and ELA Items ...	48
8.3	Summer 2007 Summary of DIF Classifications for HSAP Mathematics and ELA Items.	49
8.4	Fall 2006 HSAP Correlations among Reporting Categories.....	49
8.5	Spring 2007 HSAP Correlations among Reporting Categories .....	50
8.6	Summer 2007 HSAP Correlations among Reporting Categories .....	50

## 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 qualified for the expedited scoring process called Graduation Express. A total of 629 HSAP ELA tests and 1,469 HSAP mathematics tests were scored through this procedure.

In the summer 2007 administration, a total of 59 students took the HSAP ELA test and 245 took the HSAP mathematics. Most of these students were in adult education programs.

## **1.2 TEST DESIGN AND STRUCTURE**

Each of the pre-equated operational ELA forms consisted of 57 multiple-choice items, 2 constructed-response items, and 1 extended-response item. Each of the pre-equated operational math forms consisted of 67 multiple-choice items and 3 constructed-response items.

## **1.3 TECHNICAL REPORT CONTENT**

This technical report summarizes the results of statistical and psychometric analyses performed on the fall 2006, the spring 2007, and the summer 2007 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 2006 and summer 2007, the data summary in all chapters of this technical report includes all students who attempted the HSAP tests. For spring 2007, 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 three 2006–07 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 2006 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 2006 Summary of Student Demographics**  
**in the HSAP Sample (All Attempts)**

Demographic Category	Mathematics		ELA	
	N	%	N	%
<b>All Students</b>	12,658	100.0	9,332	100.0
<b>Gender</b>				
Female	5,969	47.2	3,868	41.4
Male	6,256	49.4	5,239	56.1
Invalid	433	3.4	225	2.4
<b>Ethnicity</b>				
African American	7,441	58.8	5,059	54.2
Asian/Pacific Islander	104	0.8	131	1.4
Hispanic	574	4.5	595	6.4
American Indian	21	0.2	17	0.2
White	3,921	31.0	3,137	33.6
Other	168	1.3	159	1.7
Invalid	405	3.2	210	2.3
<b>Grade</b>				
9	644	5.1	586	6.3
10	1,701	13.4	1,378	14.8
11	3,777	29.8	3,212	34.4
12	1,651	13.0	1,232	13.2
Invalid	4,885	38.6	2,924	31.3
<b>ESL*</b>				
No	4,265	33.69	3,407	36.5
Yes	542	4.28	615	6.6
Unknown	7,851	62.02	5,310	56.9
<b>Lunch Program</b>				
No free/reduced lunch	3,389	26.8	2,894	31.0
Free lunch	3,173	25.1	2,611	28.0
Reduced lunch	384	3.0	295	3.2
Unknown	5,712	45.1	3,532	37.8
<b>IEP**</b>				
No	7,955	62.8	5,411	58.0
Yes	4,703	37.2	3,921	42.0
Unknown	—	—	—	—
<b>Migrant</b>				
No	6,725	53.1	5,577	59.8
Yes	13	0.1	13	0.1
Unknown	5,920	46.8	3,742	40.1
<b>Attempt</b>				
1st	3,106	24.5	3,078	33.0
2nd	6,935	54.8	4,798	51.4
3rd	1,399	11.1	720	7.7
4th or more	1,218	9.6	736	7.9

\* English as a second language

\*\* individualized education program

**TABLE 2.2**  
**Spring 2007 Summary of Student Demographics**  
**in the HSAP Sample (All Attempts)**

Demographics	Mathematics		ELA	
	N	%	N	%
<b>All Students</b>	58,711	100.0	55,690	100.0
<b>Gender</b>				
Female	29,271	49.9	27,483	49.3
Male	29,167	49.7	28,006	50.3
Invalid	273	0.5	201	0.4
<b>Ethnicity</b>				
African American	24,965	42.5	22,802	40.9
Asian/Pacific Islander	668	1.1	673	1.2
Hispanic	2,074	3.5	2,051	3.7
American Indian	101	0.2	100	0.2
White	29,960	51.0	29,212	52.5
Other	596	1.0	581	1.0
Invalid	272	0.5	200	0.4
<b>Grade</b>				
9	6,641	11.3	6,563	11.8
10	47,003	80.1	46,299	83.1
11	3,381	5.8	1,984	3.6
12	1,422	2.4	648	1.2
Invalid	264	0.4	196	0.4
<b>ESL*</b>				
No	54,601	93.0	51,776	93.0
Yes	4,110	7.0	3,914	7.0
Unknown	—	—	—	—
<b>Lunch Program</b>				
No free/reduced lunch	30,649	52.2	29,500	53.0
Free lunch	23,489	40.0	21,946	39.4
Reduced lunch	4,103	7.0	3,860	6.9
<b>IEP**</b>				
No	54,601	93.0	51,776	93.0
Yes	4,110	7.0	3,914	7.0
Unknown	—	—	—	—
<b>Migrant</b>				
No	5,972	10.2	5,587	10.0
Yes	54	0.1	56	0.1
Unknown	52,685	89.7	50,047	89.9
<b>Attempt</b>				
1st	52,202	88.9	52,152	93.6
2nd	2,517	4.3	1,629	2.9
3rd	2,793	4.8	1,382	2.5
4th or more	1,199	2.0	527	0.9

\* English as a second language

\*\* individualized education program

**TABLE 2.3**  
**Summer 2007 Summary of Student Demographics**  
**in the HSAP Sample (All Attempts)**

Demographics	Mathematics		ELA	
	N	%	N	%
<b>All Students</b>	245	100.0	59	100.0
<b>Gender</b>				
Female	134	54.7	34	57.6
Male	82	33.5	24	40.7
Invalid	29	11.8	1	1.7
<b>Ethnicity</b>				
African American	176	71.8	40	67.8
Asian/Pacific Islander	1	0.4	1	1.7
Hispanic	7	2.9	8	13.6
American Indian	—	—	—	—
White	26	10.6	5	8.5
Other	7	2.9	2	3.4
Invalid	28	11.4	3	5.1
<b>Grade</b>				
9	—	—	—	—
10	2	0.8	2	3.4
11	18	7.3	—	—
12	192	78.4	55	93.2
Invalid	33	13.5	2	3.4
<b>ESL*</b>				
No	120	49.0	21	35.6
Yes	11	4.5	6	10.2
Unknown	114	46.5	32	54.2
<b>Lunch Program</b>				
No free/reduced lunch	75	30.6	14	23.7
Free lunch	91	37.1	24	40.7
Reduced lunch	12	4.9	2	3.4
Unknown	67	27.3	19	32.2
<b>IEP**</b>				
No	139	56.7	38	64.4
Yes	106	43.3	21	35.6
Unknown	—	—	—	—
<b>Migrant</b>				
No	141	57.6	42	71.2
Yes	1	0.4	1	1.7
Unknown	103	42.0	16	27.1
<b>Attempt</b>				
1st	2	0.8	—	—
2nd	6	2.4	3	5.1
3rd	19	7.8	3	5.1
4th or more	218	89.0	53	89.8

\* 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 2006a and 2007a). 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 2006 HSAP Administration (All Attempts)**

Accommodation	Mathematics		ELA	
	Regular Form (N = 12,625)	Customized Form (N = 33)	Regular Form (N = 9,301)	Customized Form (N = 31)
Setting	17.4	87.9	19.8	80.6
Presentation	—	—	—	67.7
Timing	1.7	66.7	2.1	61.3
Schedule	0.8	63.6	1.0	61.3
Response options	0.5	69.7	1.3	64.5
Loose-leaf	0.3	—	0.0	—
Large-print	0.1	—	0.1	—
Spelling dictionary	—	—	1.8	45.2
Audiocassette	5.5	3.0	—	—
Oral script	6.7	9.1	—	—
Signed administration	—	75.8	—	—
Braille	—	24.2	—	16.1
Other	0.1	9.1	0.0	—

**TABLE 2.5**  
**Percentages of Students with Accommodations in the**  
**Spring 2007 HSAP Administration (All Attempts)**

Accommodation	Mathematics		ELA	
	Regular Form (N = 58,650)	Customized Form (N = 61)	Regular Form (N = 55,627)	Customized Form (N = 63)
Setting	7.9	83.6	8.3	85.7
Presentation	—	—	0.0	63.5
Timing	0.5	60.7	0.5	60.3
Schedule	0.2	57.4	0.3	50.8
Response options	0.2	54.1	0.5	55.6
Loose-leaf	0.0	—	0.0	—
Large-print	0.1	—	0.0	—
Spelling dictionary	—	—	0.6	42.9
Audiocassette	2.6	4.9	—	—
Oral script	2.6	1.6	—	—
Signed administration	—	90.2	—	—
Braille	—	9.8	0.1	7.9
Other	0.0	6.6	0.0	—

**TABLE 2.6**  
**Percentages of Students with Accommodations in the**  
**Summer 2007 Accommodations (All Attempts)**

Accommodation	Mathematics		ELA	
	Regular Form (N = 243)	Customized Form (N = 2)	Regular Form (N = 57)	Customized Form (N = 2)
Setting	15.2	100.0	26.3	100.0
Presentation	—	—	—	—
Timing	3.7	50.0	8.8	50.0
Schedule	—	—	—	—
Response options	1.2	50.0	7.0	50.0
loose-leaf	0.4	—	1.8	—
Large-print	—	—	—	—
Spelling dictionary	—	—	—	7.0
Audiocassette	7.4	—	—	—
Oral script	7.8	—	—	—
Signed administration	—	100.0	—	—
Braille	—	—	—	—
Other	—	—	5.3	—

### 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 2006–07 HSAP Administrations (All Attempts)**

<b>Modification</b>	<b>Regular Form</b>	<b>Customized Form</b>
<b>Fall 2006</b>	<b>(N = 9,301 )</b>	<b>(N = 31)</b>
Alternative scoring	7.3	64.5
Extended writing option	1.9	45.2
Audiocassette	7.7	—
Oral administration	7.9	3.2
Signed administration	—	83.9
<b>Spring 2007</b>	<b>Regular Form</b>	<b>Customized Form</b>
	<b>(N = 55,627)</b>	<b>(N = 63)</b>
Alternative scoring	2.3	73.0
Extended writing option	0.5	41.3
Audiocassette	2.8	1.6
Oral administration	2.7	1.6
Signed administration	—	92.1
<b>Summer 2007</b>	<b>Regular Form</b>	<b>Customized Form</b>
	<b>(N = 57)</b>	<b>(N = 2)</b>
Alternative scoring	22.8	50.0
Extended writing option	10.5	50.0
Audiocassette	15.8	—
Oral administration	22.8	—
Signed administration	—	—

### 2.3 STUDENT QUESTIONNAIRES

After the test administration, students were instructed to answer the questions on the HSAP student questionnaire: for mathematics, there were 12 questions in Fall 2006 and 17 in Spring 2007 and Summer 2007; for ELA, there were 8 questions in Fall 2006, Spring 2007, and Summer 2007. The questionnaire topics encompassed test difficulty, classroom activities, and (for mathematics only) calculator use.

## Chapter 3

### TEST ADMINISTRATION

#### 3.1 TEST ADMINISTRATION WINDOW

The HSAP ELA operational tests for Fall 2006 and Spring 2007 were conducted in two sessions over two days. The mathematics tests were conducted in one day. For the Summer 2007 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 2007 administration.

**TABLE 3.1**  
**2006–07 HSAP Test Administration Schedule**

Fall 2006		Spring and Summer 2007	
Date	Test	Date	Test
October 24	ELA (day 1)	April 24	ELA (day 1)
October 25	ELA (day 2)	April 25	ELA (day 2)
October 26	Mathematics	April 26	Mathematics
October 27–November 3	Makeup tests window	April 27–May 4	Makeup tests window
		July 17–19	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 2006, Spring 2007, and Summer 2007 (All Attempts)**

Time Taken	Fall 2006			Spring 2007			Summer 2007		
	Math % (N = 12,658)	ELA % (N = 9,332)		Math % (N = 58,711)	ELA % (N = 55,690)		Math % (N = 245)	ELA % (N = 59)	
		Session 1	Session 2		Session 1	Session 2		Session 1	Session 2
15 min	0.5	1.8	0.8	0.4	0.9	0.6	—	1.7	—
30 min	1.8	9.0	3.2	0.9	6.8	4.9	—	8.5	—
45 min	6.4	18.3	9.3	4.0	19.2	20.1	—	22.0	—
1 hr	13.6	19.6	15.0	13.8	25.8	27.9	0.8	15.3	1.7
1 hr 15 min	15.4	15.4	15.9	19.8	18.9	19.3	4.9	8.5	10.2
1 hr 30	15.3	10.4	13.0	19.1	11.6	10.8	8.6	6.8	10.2
1 hr 45	11.2	7.2	11.0	14.2	6.6	5.8	13.1	10.2	16.9
2 hr	9.8	4.7	8.8	10.3	3.8	3.5	11.4	5.1	16.9
2 hr 15 min	6.3	3.4	5.8	5.8	1.7	1.9	12.2	3.4	8.5
2 hr 30 min	4.6	1.5	3.8	3.5	1.0	1.0	8.6	15.3	6.8
2 hr 45 min	2.9	1.3	2.3	1.9	0.5	0.5	10.2	3.4	—
3 hr +	5.1	1.8	4.1	3.2	0.9	1.0	22.9	1.7	23.7
Invalid	6.9	5.5	6.9	2.9	2.3	2.8	7.3	8.5	5.1

### 3.3 ADMINISTRATION MANUALS

Working with SCDE staff, Pearson Educational Measurement (PEM) staff drafted the administration manuals for the test. SCDE staff reviewed and revised the manuals, and the PEM 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 2006b and 2007b). 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 2007 supplement (SCDE 2007c) was produced to provide specific information for the Summer 2007 administration and was used in conjunction with the information provided in the Spring 2007 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 PEM staff trained the district staff in attendance.

PEM 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 the PEM 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 PEM shipment to each district office also included a 10 percent overage of all test materials—with the exception of customized formats, which were sent only in the quantities ordered. The 10 percent overage was in addition to the 5 percent overage included in school shipments. Overage materials for the districts were to be used by the DTCs to fulfill any additional materials requests from the STCs.

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

With the PEM shipment of overage materials, the DTCs 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 PEM. 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, PEM sent missing materials letters to districts identifying the number of unreturned secure materials and the barcode numbers of each missing document. The districts had two weeks to respond to the letter before PEM 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**

<b>Rule</b>	<b>First Score (R1)</b>	<b>Second Score (R2)</b>	<b>Action</b>	<b>Back Reading (BR)</b>	<b>Resolution Score (RS) [Third Score]</b>	<b>Final Score (F)</b>
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**

<b>Rule</b>	<b>Supervisor First Score (S1)</b>	<b>Supervisor Second Score (S2)</b>	<b>Action</b>	<b>BR</b>	<b>Supervisor Resolution Score (S3)</b>	<b>Final Score (F)</b>
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**

<b>Fall 2006, Spring and Summer 2007</b>	<b>Algebra</b>	<b>Data Analysis and Probability</b>	<b>Measurement and Geometry</b>	<b>Number and Operations</b>	<b>Integrated Responses</b>
Percentage	27%	11%	27%	23%	13%
Multiple-choice points	19	8	19	6	—
Constructed-response points	—	—	—	—	9

**TABLE 4.4**

**HSAP ELA: Distribution of Score Point Values by Reporting Category**

	<b>Reading Process and Comprehension</b>	<b>Analysis of Texts</b>	<b>Word Study and Analysis</b>	<b>Research</b>	<b>Writing</b>
<b>Fall 2006</b>					
Percentage	24%	20%	8%	8%	40%
Multiple-choice points	20	16	8	8	8
Constructed-response points	3	3	—	—	—
Extended-response points	—	—	—	—	30
<b>Spring and Summer 2007</b>					
Percentage	24%	20%	8%	8%	40%
Multiple-choice points	20	16	8	8	8
Constructed-response points	3	3	—	—	—
Extended-response points	—	—	—	—	30

### **4.3 SCORING PROCESS**

PEM scored all items: multiple-choice items were scored by PEM’s electronic scanning system; constructed-response (CR) and extended-response (ER) items were scored by trained personnel using the ePEN (Electronic Performance Evaluation Network) system. There were two scoring sites for the HSAP in ELA: the Fall 2006 and Summer 2007 tests were scored in Iowa City, Iowa; the Spring 2007 test in Mesa, Arizona. There were also two scoring sites for the HSAP in mathematics: the Fall 2006 and Spring 2007 tests were scored in Lansing, Michigan; the Summer 2007 test in Iowa City, Iowa.

Prior to actual scoring of the constructed-response and extended-response items, range-finding meetings were held in Columbia, South Carolina: September 6–8, 2006, for the Fall 2006 administration; January 22–24, 2007, for the Spring 2007 administration; and February 26–28, 2007, for the Summer 2007 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 2007 and summer 2007 training sessions. Throughout the scoring process, PEM staff posted the performance of each reader (reader-reliability statistics) once a day on PEM’s SchoolHouse Web site for 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
<b>Mathematics</b>			
<b>Fall 2006</b>			
CR1	1,896	74.4	98.8
CR2	1,893	82.7	99.4
CR3	1,893	80.7	99.5
<b>Spring 2007</b>			
CR1	8,590	80.7	99.6
CR2	8,578	87.3	99.5
CR3	8,585	76.8	99.2
<b>Summer 2007</b>			
CR1	37	91.9	100.0
CR2	40	100.0	100.0
CR3	38	97.4	100.0
<b>ELA</b>			
<b>Fall 2006</b>			
CR1	1,407	68.4	98.3
CR2	1,413	68.0	95.6
ER content and development	9,332	63.2	98.2
ER organization	9,332	59.9	80.7
ER voice	9,332	62.5	99.2
ER convention	9,332	53.6	94.6
<b>Spring 2007</b>			
CR1	8,253	81.7	99.6
CR2	8,254	75.7	99.6
ER content and development	55,050	72.4	99.2
ER organization	55,050	69.9	84.8
ER voice	55,050	71.9	99.5
ER convention	55,050	73.8	97.3
<b>Summer 2007</b>			
CR1	8	87.5	100.0
CR2	8	100.0	100.0
ER content and development	59	76.3	100.0
ER organization	59	82.7	88.1
ER voice	59	83.1	100.0
ER convention	59	54.2	100.0

#### **4.5 TESTED/NOT TESTED FLAG**

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

## Chapter 5

### TECHNICAL CHARACTERISTICS OF ITEMS

This chapter reports the results of item analyses based on classical test theory (CTT) using a proprietary program designed by the PEM. 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 PEM staff decided to exclude from the CTT item analyses and item calibrations those students who had used customized materials and those who had received the alternative scoring rubric modification.

#### 5.1 ITEM NONRESPONSE RATES

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

The percentage of students who responded to the last two items on a given test form was computed. Table 5.1, 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 14 in fall and item 21 in spring in session 1 and item 59 in fall, and item 73 in spring in session 2 for ELA. Item nonresponse rates were computed for each ELA session separately. Students tend to leave CR items blank more often than they leave MC items blank, especially when the CR items appear at the end of the test.

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

Subject	Last Item	Second-to-Last Item
<b>Fall 2006</b>		
Mathematics	92.0 (CR)	93.8 (CR)
ELA Session 1	98.5 (MC)	98.8 (MC)
ELA Session 2	98.1 (MC)	98.4 (MC)
<b>Spring 2007</b>		
Mathematics	91.3 (CR)	96.8 (CR)
ELA Session 1	96.2 (MC)	97.3 (MC)
ELA Session 2	97.2 (MC)	97.4 (MC)
<b>Summer 2007</b>		
Mathematics	94.3 (CR)	99.6 (CR)
ELA Session 1	98.3 (MC)	98.3 (MC)
ELA Session 2	98.3 (MC)	98.3 (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 2006	Spring 2007	Summer 2007			
<b>Mathematics</b>									
Multiple-choice	62	0.52	0.32	62	0.66	0.43	62	0.51	0.16
Constructed-response	3	0.28	0.63	3	0.55	0.70	3	0.47	0.79
Number and Operations	16	0.60	0.33	16	0.72	0.41	16	0.63	0.17
Algebra	19	0.53	0.31	19	0.68	0.42	19	0.51	0.12
Measurement and Geometry	19	0.43	0.30	19	0.59	0.45	19	0.38	0.12
Data Analysis and Probability	8	0.51	0.33	8	0.66	0.48	8	0.58	0.20
<b>ELA</b>									
Multiple-choice	60	0.57	0.31	60	0.73	0.35	60	0.49	0.21
Constructed-response	2	0.34	0.45	2	0.46	0.53	2	0.41	0.26
Extended-response	1	0.78	0.56	1	0.89	0.66	1	0.74	0.40
Reading Process and Comprehension	21	0.56	0.38	21	0.69	0.37	21	0.50	0.20
Analysis of Texts	17	0.58	0.37	17	0.71	0.40	17	0.51	0.26
Word Study and Analysis	9	0.60	0.30	8	0.76	0.26	8	0.47	0.24
Research	7	0.73	0.28	8	0.86	0.34	8	0.45	0.18
Writing	9	0.43	0.36	9	0.63	0.41	9	0.68	0.18

## 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 PEM 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 ( $\theta_c$ ) 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 PEM 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 PEM, 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 PEM 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
<b>4</b>	<p>The Level 4 student</p> <ul style="list-style-type: none"> <li>• has demonstrated an exceptional command of skills and knowledge required of high school students in South Carolina</li> <li>• analyzes, evaluates, and/or synthesizes mathematical concepts and procedures and solves problems using advanced arithmetic, algebraic, and measurement/geometric concepts and relationships</li> <li>• analyzes data representations and applies probability concepts</li> <li>• supports answers with mathematical work and/or explanations that thoroughly communicate mathematical reasoning</li> <li>• has met the exit examination requirement for a South Carolina high school diploma</li> </ul>
<b>3</b>	<p>The Level 3 student</p> <ul style="list-style-type: none"> <li>• has demonstrated proficiency in skills and knowledge required of high school students in South Carolina</li> <li>• applies mathematical concepts and procedures and solves problems using arithmetic, algebraic, and measurement/geometric concepts and relationships</li> <li>• interprets data representations and demonstrates a knowledge of probability concepts</li> <li>• supports answers with mathematical work and/or explanations that clearly communicate mathematical reasoning</li> <li>• has met the exit examination requirement for a South Carolina high school diploma</li> </ul>
<b>2</b>	<p>The Level 2 student</p> <ul style="list-style-type: none"> <li>• has demonstrated competence in skills and knowledge required of high school students in South Carolina</li> <li>• demonstrates an acceptable knowledge of fundamental mathematical concepts and procedures and solves problems using essential arithmetic, algebraic, and measurement/geometric concepts and relationships</li> <li>• demonstrates a knowledge of basic data representations and probability concepts</li> <li>• supports answers with mathematical work and/or explanations that adequately communicate mathematical reasoning</li> <li>• has met the exit examination requirement for a South Carolina high school diploma</li> </ul>
<b>1</b>	<p>The Level 1 student</p> <ul style="list-style-type: none"> <li>• has not demonstrated competence in the skills and knowledge required of high school students in South Carolina</li> <li>• demonstrates a limited understanding of mathematical concepts</li> <li>• is able to use arithmetic, algebraic, and measurement/geometric concepts and relationships</li> <li>• demonstrates a knowledge of simple data representations and probability concepts</li> <li>• supports answers with mathematical work and/or explanations that minimally communicate mathematical reasoning</li> <li>• has not met the exit examination requirement for a South Carolina high school diploma</li> </ul>

**TABLE 6.2**  
**Description of Achievement Levels for the HSAP ELA Test**

<b>Level</b>	<b>Description</b>
<b>4</b>	<p>The Level 4 student</p> <ul style="list-style-type: none"> <li>• has demonstrated an exceptional command of skills and knowledge required of high school students in South Carolina</li> <li>• demonstrates comprehension of complex ideas and connects those ideas within a text, across texts, and beyond the text</li> <li>• displays exceptional writing skills by engaging the reader, effectively developing and organizing ideas, and using relevant supporting details, vivid language, and Standard American English</li> <li>• has met the exit examination requirement for a South Carolina high school diploma</li> </ul>
<b>3</b>	<p>The Level 3 student</p> <ul style="list-style-type: none"> <li>• has demonstrated proficiency in skills and knowledge required of high school students in South Carolina</li> <li>• demonstrates comprehension of complex ideas and connects those ideas within a text and across texts</li> <li>• displays effective writing skills by sustaining the reader’s interest, clearly developing and organizing ideas, and using relevant supporting details and Standard American English</li> <li>• has met the exit examination requirement for a South Carolina high school diploma</li> </ul>
<b>2</b>	<p>The Level 2 student</p> <ul style="list-style-type: none"> <li>• has demonstrated competence in skills and knowledge required of high school students in South Carolina</li> <li>• demonstrates comprehension of essential ideas and shows some logical connections of those ideas within a text</li> <li>• displays acceptable writing skills by showing some awareness of audience, developing and organizing ideas, and using relevant supporting details and Standard American English</li> <li>• has met the exit examination requirement for a South Carolina high school diploma</li> </ul>
<b>1</b>	<p>The Level 1 student</p> <ul style="list-style-type: none"> <li>• has not demonstrated competence in skills and knowledge required of high school students in South Carolina</li> <li>• demonstrates limited comprehension of ideas and tenuous connections of those ideas within a text</li> <li>• 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</li> <li>• has not met the exit examination requirement for a South Carolina high school diploma</li> </ul>

## 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**

	<b>Level 2</b>	<b>Level 3</b>	<b>Level 4</b>
<b>Mathematics</b>			
Rasch Ability	-0.224	0.658	1.584
Scale Score	200	220	241
<b>ELA</b>			
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 ( $\theta_c$ ) was computed as cut score ( $\theta_c$ )  $\pm$  1 SE( $\theta_c$ ). 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 ( $\theta$ )	SE( $\theta$ )	68% Confidence Interval	
			$\theta - 1SE$	$\theta + 1SE$
<b>Mathematics</b>				
<b>Fall 2006</b>				
Number and Operations	0.034	0.575	-0.541	0.609
Algebra	-0.001	0.489	-0.490	0.488
Measurement and Geometry	-0.065	0.473	-0.538	0.408
Data Analysis and Probability	-0.094	0.722	-0.816	0.628
<b>Spring 2007</b>				
Number and Operations	0.028	0.559	-0.531	0.587
Algebra	-0.028	0.479	-0.507	0.451
Measurement and Geometry	-0.041	0.472	-0.513	0.431
Data Analysis and Probability	-0.061	0.740	-0.801	0.679
<b>Summer 2007</b>				
Number and Operations	0.044	0.574	-0.530	0.618
Algebra	-0.038	0.478	-0.516	0.440
Measurement and Geometry	-0.056	0.476	-0.532	0.420
Data Analysis and Probability	-0.165	0.730	-0.895	0.565
<b>ELA</b>				
<b>Fall 2006</b>				
Reading Process and Comprehension	0.145	0.450	-0.305	0.595
Analysis of Texts	0.091	0.485	-0.294	0.576
Word Study and Analysis	0.512	0.761	-0.249	1.273
Research	0.107	0.802	-0.695	0.909
Writing	0.024	0.421	-0.397	0.445
<b>Spring 2007</b>				
Reading Process and Comprehension	0.052	0.446	-0.394	0.498
Analysis of Texts	0.156	0.481	-0.325	0.637
Word Study and Analysis	0.045	0.754	-0.709	0.799
Research	0.472	0.748	-0.276	1.220
Writing	0.043	0.395	-0.352	0.438
<b>Summer 2007</b>				
Reading Process and Comprehension	0.088	0.444	-0.356	0.532
Analysis of Texts	0.028	0.503	-0.475	0.531
Word Study and Analysis	0.030	0.767	-0.737	0.797
Research	0.197	0.724	0.527	0.921
Writing	0.028	0.386	-0.358	0.414

## 6.9 PERCENTAGE OF STUDENTS IN EACH ACHIEVEMENT LEVEL

Tables 6.5 through 6.10, 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 6.11 through 6.16 provide the information for content areas. The information is summarized for Level 1 and at or above Level 2 for all students by gender and by ethnic group. Of those students who took both the mathematics and the ELA tests for the first time, 74 percent in Spring 2007 passed both tests. In Fall 2006, of those students who took both tests, 35 percent passed both tests.

**TABLE 6.5**  
**Fall 2006 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	
<b>Overall</b>	60.0	28.7	6.6	4.7	40.0	11.3	12,620
<b>Gender</b>							
Female	58.2	30.9	6.6	4.2	41.7	10.8	5,952
Male	39.5	30.4	16.1	14.0	60.5	30.1	428
Invalid	63.1	26.4	5.9	4.6	36.9	10.5	6,240
<b>Ethnicity</b>							
African American	69.1	27.9	2.3	0.7	30.9	3.0	7,421
Asian/Pacific Islander	57.1	14.3	23.8	4.8	42.9	28.6	21
Hispanic	41.3	22.1	18.3	18.3	58.7	36.6	104
American Indian	62.6	26.3	6.7	4.4	37.4	11.1	570
White	40.5	26.2	14.9	18.5	59.6	33.4	168
Other	33.8	31.5	18.3	16.5	66.3	34.8	400
Unknown	46.4	30.6	12.7	10.3	53.6	23.0	3,912
<b>ESL*</b>							
Yes	58.0	29.8	8.0	4.3	35.4	10.2	540
No	52.8	28.6	10.7	7.9	47.2	18.6	4,263
Unknown	64.1	28.6	4.3	3.1	36.0	7.4	7,817
<b>Lunch Program</b>							
No free/reduced lunch	43.3	29.1	15.0	12.6	56.7	27.6	3,377
Free lunch	66.9	27.0	4.3	1.7	33.0	6.0	3,167
Reduced lunch	56.5	33.9	6.3	3.4	43.6	9.7	384
Unknown	66.3	29.0	2.9	1.9	33.8	4.8	5,692
<b>IEP**</b>							
Yes	44.3	32.7	13.3	9.8	55.8	23.1	4,694
No	69.3	26.3	2.6	1.8	30.7	4.4	7,926
Unknown	—	—	—	—	—	—	—
<b>Migrant</b>							
Yes	69.2	30.8	—	—	—	—	13
No	55.0	27.9	9.8	7.3	45.0	17.1	6,709

\* English as a second language

\*\* individualized education program

**TABLE 6.6**  
**Spring 2007 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	
<b>Overall</b>	20.8	27.3	27.0	24.9	79.2	51.9	52,116
<b>Gender</b>							
Female	18.6	29.4	27.9	24.1	81.4	52.0	26,106
Male	22.8	25.1	26.2	25.9	77.2	52.1	25,813
Invalid	50.3	24.4	15.7	9.6	49.7	25.3	197
<b>Ethnicity</b>							
African American	33.1	34.3	22.5	10.0	66.8	32.5	20,443
Asian/Pacific Islander	10.1	17.4	22.7	49.8	89.9	72.5	621
Hispanic	29.2	28.4	26.1	16.2	70.7	42.3	1,771
American Indian	17.2	28.0	28.0	26.9	82.9	54.9	93
White	11.5	22.3	30.5	35.8	88.6	66.3	28,398
Other	20.5	29.7	27.4	22.4	79.5	49.8	532
Unknown	50.8	22.8	16.4	10.1	49.3	26.5	189
<b>ESL*</b>							
Yes	36.8	28.6	21.3	8.7	37.9	23.3	1,270
No	19.6	30.4	28.5	21.5	80.4	50.0	46,403
Unknown	42.2	29.2	17.7	10.9	57.8	28.5	883
<b>Lunch Program</b>							
No free/reduced lunch	12.0	22.4	29.9	35.7	88.0	65.6	28,464
Free lunch	33.2	33.4	22.8	10.6	66.8	33.4	19,686
Reduced lunch	19.6	32.3	28.4	19.8	80.5	48.2	3,624
Unknown	51.2	26.6	13.7	8.5	48.8	22.2	342
<b>IEP**</b>							
Yes	28.1	31.7	23.9	16.3	71.9	40.2	3,557
No	20.0	27.2	27.5	25.6	80.3	53.1	49,643
Unknown	49.5	28.7	14.3	7.5	50.5	21.8	1,203
<b>Migrant</b>							
Yes	34.1	20.5	27.3	18.2	66.0	45.5	44
No	40.3	28.1	19.6	12.0	59.7	31.6	4,818
Unknown	18.8	27.2	27.8	26.2	81.2	54.0	47,254

\* English as a second language

\*\* individualized education program

**TABLE 6.7**  
**Summer 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	
<b>Overall</b>	55.1	43.7	0.8	0.4	44.9	1.2	245
<b>Gender</b>							
Female	53.7	44.0	1.5	0.7	46.2	2.2	134
Male	62.2	37.8	—	—	—	—	82
Invalid	41.4	58.6	—	—	—	—	29
<b>Ethnicity</b>							
African American	56.8	42.0	0.6	0.6	43.2	1.2	176
Asian/Pacific Islander	100.0	—	—	—	—	—	0
Hispanic	57.1	42.9	—	—	—	—	1
American Indian	57.1	42.9	—	—	—	—	7
White	42.9	57.1	—	—	—	—	7
Other	53.8	42.3	3.8	—	—	—	28
Unknown	100.0	—	—	—	—	—	26
<b>ESL*</b>							
Yes	72.3	27.3	—	—	—	—	11
No	53.3	45.0	0.9	0.9	45.0	1.7	120
Unknown	57.0	42.1	0.9	—	—	—	114
<b>Lunch Program</b>							
No free/reduced lunch	58.7	41.3	—	—	—	—	75
Free lunch	58.2	40.7	1.1	—	—	—	91
Reduced lunch	66.7	33.3	—	—	—	—	12
Unknown	44.8	52.2	3.0	—	—	—	67
<b>IEP**</b>							
Yes	67.9	30.9	—	—	—	—	81
No	50.9	48.1	0.9	—	—	—	106
Unknown	44.8	53.5	1.7	—	—	—	58
<b>Migrant</b>							
Yes	100.0	—	—	—	—	—	1
No	60.3	39.0	0.7	—	—	—	141
Unknown	48.5	49.5	1.9	—	—	—	103

\* English as a second language

\*\* individualized education program

**TABLE 6.8**  
**Fall 2006 HSAP ELA 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	
<b>Overall</b>	43.0	35.6	11.9	9.4	56.9	21.3	9,305
<b>Gender</b>							
Female	36.6	38.4	13.0	12.0	63.4	25.0	3,859
Male	48.2	33.8	10.8	7.2	51.8	18.0	5,226
Invalid	32.3	30.0	20.5	17.3	67.8	37.8	220
<b>Ethnicity</b>							
African American	52.6	39.0	6.1	2.3	47.4	8.4	5,044
Asian/Pacific Islander	39.7	35.9	9.2	15.3	60.4	24.5	131
Hispanic	53.5	29.1	11.8	5.6	46.5	17.4	594
American Indian	47.1	35.3	11.8	5.9	53.0	17.7	17
White	27.8	31.9	20.2	20.1	72.2	40.3	3,131
Other	27.8	29.1	22.8	20.3	72.2	43.1	158
Unknown	26.7	32.0	21.8	19.4	73.2	41.2	206
<b>ESL*</b>							
Yes	52.4	30.9	10.5	6.2	31.6	14.6	615
No	33.3	33.8	17.8	15.1	66.7	33.0	3,406
Unknown	48.3	37.3	8.3	6.2	51.8	14.5	5,284
<b>Lunch Program</b>							
No free/reduced lunch	26.0	28.8	21.9	23.3	74.0	45.2	2,892
Free lunch	51.3	37.4	8.6	2.7	48.7	11.3	2,609
Reduced lunch	38.6	38.3	14.2	8.8	61.3	23.0	295
Unknown	51.3	39.7	6.0	3.0	48.7	9.0	3,509
<b>IEP**</b>							
Yes	26.2	33.8	21.0	19.0	73.8	40.0	3,918
No	55.2	36.9	5.3	2.5	44.7	7.8	5,387
Unknown	—	—	—	—	—	—	—
<b>Migrant</b>							
Yes	76.9	15.4	7.7	—	—	—	13
No	37.4	33.1	15.8	13.7	62.6	29.5	5,573
Unknown	51.3	39.4	6.2	3.2	48.8	9.4	3,719

\* English as a standard language

\*\*individualized education program

**TABLE 6.9**  
**Spring 2007 HSAP ELA 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	
<b>Overall</b>	12.1	28.3	32.0	27.6	87.9	59.6	52,075
<b>Gender</b>							
Female	8.4	27.5	33.2	30.9	91.6	64.1	26,099
Male	15.8	29.0	30.9	24.3	84.2	55.2	25,812
Invalid	26.2	33.5	27.4	12.8	73.7	40.2	164
<b>Ethnicity</b>							
African American	18.6	39.6	30.0	11.8	81.4	41.8	20,407
Asian/Pacific Islander	11.9	20.2	29.4	38.5	88.1	67.9	620
Hispanic	23.9	33.7	27.1	15.3	76.1	42.4	1,772
American Indian	16.8	23.2	25.3	34.7	83.2	60.0	95
White	6.6	20.0	33.9	39.5	93.4	73.4	28,419
Other	14.5	27.2	31.7	26.6	85.5	58.3	537
Unknown	28.0	32.5	25.5	14.0	72.0	39.5	157
<b>ESL*</b>							
Yes	35.2	35.6	20.3	8.9	62.5	28.7	1,277
No	11.1	27.9	32.6	28.5	89.0	61.1	49,626
Unknown	31.4	36.9	21.9	9.8	68.6	31.7	1,172
<b>Lunch Program</b>							
No free/reduced lunch	6.0	20.1	34.4	39.6	94.1	74.0	28,454
Free lunch	21.0	39.3	27.9	11.8	79.0	39.7	19,704
Reduced lunch	11.1	31.7	36.8	20.5	89.0	57.3	3,610
Unknown	28.0	35.5	23.5	13.0	72.0	36.5	307
<b>IEP**</b>							
Yes	13.8	36.1	30.8	19.2	86.1	50.0	3,595
No	12.0	27.7	32.1	28.2	88.0	60.3	48,480
Unknown	—	—	—	—	—	—	—
<b>Migrant</b>							
Yes	31.0	33.3	23.8	11.9	69.0	35.7	42
No	27.2	33.9	25.1	13.8	72.8	38.9	4,862
Unknown	10.5	27.7	32.8	29.0	89.5	61.8	47,171

\* English as a second language

\*\* individualized education program

**TABLE 6.10**  
**Summer 2007 HSAP ELA 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	
<b>Overall</b>	72.9	25.4	1.7	—	—	—	59
<b>Gender</b>							
Female	76.5	20.6	2.9	—	—	—	34
Male	66.7	33.3	—	—	—	—	24
Invalid	100.0	—	—	—	—	—	1
<b>Ethnicity</b>							
African American	72.5	25.0	2.5	—	—	—	40
Asian/Pacific Islander	100.0	—	—	—	—	—	1
Hispanic	62.5	37.5	—	—	—	—	8
American Indian	—	—	—	—	—	—	0
White	80.0	20.0	—	—	—	—	5
Other	50.0	50.0	—	—	—	—	2
Unknown	100.0	—	—	—	—	—	3
<b>ESL*</b>							
Yes	83.3	16.7	—	—	—	—	6
No	85.7	14.3	—	—	—	—	21
Unknown	68.8	28.1	3.1	—	—	—	32
<b>Lunch Program</b>							
No free/reduced lunch	78.6	14.3	7.1	—	—	—	14
Free lunch	79.2	20.8	—	—	—	—	24
Reduced lunch	50.0	50.0	—	—	—	—	2
Unknown	63.2	36.8	—	—	—	—	19
<b>IEP**</b>							
Yes	84.6	11.5	3.9	—	—	—	26
No	61.9	38.1	—	—	—	—	21
Unknown	66.7	33.3	—	—	—	—	12
<b>Migrant</b>							
Yes	100.0	—	—	—	—	—	1
No	76.2	21.4	2.4	—	—	—	42
Unknown	62.5	37.5	—	—	37.5	0.0	16

\* English as a second language

\*\* individualized education program

**TABLE 6.11**  
**Fall 2006 HSAP Mathematics Operational Test:**  
**Content-Area Information (All Attempts)**

<b>Subgroup</b>	<b>Level 1</b>				<b>Level 2 and Above</b>			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<b>N1*</b>	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<b>N2**</b>
<b>Number and Operations</b>								
All students	72.5%	25.9%	1.5%	7,571	11.6%	46.5%	41.9%	5,049
Females	73.4%	25.4%	1.2%	3,466	12.9%	48.7%	38.4%	2,486
Males	71.6%	26.7%	1.7%	3,936	10.8%	45.5%	43.7%	2,304
African Americans	73.8%	24.9%	1.3%	5,131	16.5%	56.9%	26.7%	2,290
Whites	69.9%	28.0%	2.1%	1,817	8.0%	37.7%	54.4%	2,095
<b>Algebra</b>								
All students	54.5%	42.6%	2.9%	7,571	3.3%	43.5%	53.2%	5,049
Females	51.1%	45.2%	3.7%	3,466	3.2%	44.7%	52.2%	2,486
Males	57.1%	40.6%	2.3%	3,936	3.5%	43.8%	52.6%	2,304
African Americans	54.2%	42.6%	3.1%	5,131	4.1%	54.2%	41.7%	2,290
Whites	54.8%	42.8%	2.5%	1,817	2.7%	36.2%	61.1%	2,095
<b>Measurement and Geometry</b>								
All students	55.3%	44.1%	0.6%	7,571	6.2%	55.9%	38.0%	5,049
Females	53.4%	45.9%	0.7%	3,466	6.4%	58.4%	35.2%	2,486
Males	57.0%	42.5%	0.6%	3,936	6.3%	55.1%	38.6%	2,304
African Americans	56.4%	42.9%	0.6%	5,131	8.3%	71.5%	20.1%	2,290
Whites	51.8%	47.6%	0.6%	1,817	4.3%	43.4%	52.3%	2,095
<b>Data Analysis and Probability</b>								
All students	35.2%	58.8%	6.0%	7,571	2.3%	45.5%	52.2%	5,049
Females	32.5%	61.2%	6.3%	3,466	2.5%	46.6%	50.8%	2,486
Males	37.2%	57.0%	5.8%	3,936	2.2%	45.6%	52.2%	2,304
African Americans	35.8%	58.3%	5.9%	5,131	2.8%	55.2%	42.1%	2,290
Whites	35.2%	58.8%	6.0%	7,571	2.3%	45.5%	52.2%	5,049

\* total number of students in Level 1

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

**TABLE 6.12**  
**Spring 2007 HSAP Mathematics Operational Test:**  
**Content-Area Information (First Attempt)**

<b>Subgroup</b>	<b>Level 1</b>				<b>Level 2 and Above</b>			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<b>N1*</b>	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<b>N2**</b>
<b>Number and Operations</b>								
All students	79.7%	19.4%	0.9%	10,848	6.7%	30.0%	63.3%	41,268
Females	79.8%	19.6%	0.6%	4,858	7.3%	31.2%	61.5%	21,248
Males	79.5%	19.3%	1.2%	5,891	6.1%	28.6%	65.3%	19,922
African Americans	80.8%	18.4%	0.9%	6,775	9.9%	40.0%	50.1%	13,668
Whites	76.8%	22.0%	1.2%	3,253	4.9%	24.5%	70.7%	25,145
<b>Algebra</b>								
All students	59.4%	37.6%	3.0%	10,848	1.4%	19.5%	79.2%	41,268
Females	53.3%	43.0%	3.7%	4,858	1.0%	18.6%	80.3%	21,248
Males	64.3%	33.2%	2.5%	5,891	1.7%	20.3%	78.0%	19,922
African Americans	59.0%	38.1%	3.0%	6,775	1.6%	26.0%	72.4%	13,668
Whites	60.6%	36.7%	2.7%	3,253	1.2%	16.0%	82.8%	25,145
<b>Measurement and Geometry</b>								
All students	60.4%	39.0%	0.6%	10,848	2.3%	31.5%	66.2%	41,268
Females	60.0%	39.6%	0.4%	4,858	3.0%	34.4%	62.6%	21,248
Males	60.7%	38.5%	0.7%	5,891	1.6%	28.3%	70.1%	19,922
African Americans	62.5%	37.1%	0.5%	6,775	4.2%	47.5%	48.2%	13,668
Whites	56.0%	43.0%	0.9%	3,253	1.3%	23.0%	75.7%	25,145
<b>Data Analysis and Probability</b>								
All students	43.2%	53.9%	2.9%	10,848	1.3%	27.6%	71.1%	41,268
Females	42.3%	54.9%	2.8%	4,858	1.4%	29.0%	69.5%	21,248
Males	44.1%	53.0%	2.9%	5,891	1.1%	25.9%	72.9%	19,922
African Americans	45.2%	52.5%	2.4%	6,775	2.1%	39.0%	58.9%	13,668
Whites	39.0%	56.7%	4.2%	3,253	0.8%	20.9%	78.3%	25,145

\* total number of students in Level 1

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

**TABLE 6.13**  
**Summer 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**
<b>Number and Operations</b>								
All students	60.0%	37.0%	3.0%	135	13.6%	63.6%	22.7%	110
Females	59.7%	37.5%	2.8%	72	12.9%	61.3%	25.8%	62
Males	54.9%	41.2%	3.9%	51	16.1%	61.3%	22.6%	31
African Americans	60.0%	37.0%	3.0%	100	11.8%	63.2%	25.0%	76
Whites	50.0%	50.0%	—	14	25.0%	58.3%	16.7%	12
<b>Algebra</b>								
All students	52.6%	43.7%	3.7%	135	5.5%	71.8%	22.7%	110
Females	50.0%	48.6%	1.4%	72	6.5%	62.9%	30.6%	62
Males	60.8%	33.3%	5.9%	51	3.2%	87.1%	9.7%	31
African Americans	52.0%	44.0%	4.0%	100	2.6%	69.7%	27.6%	76
Whites	64.3%	35.7%	—	14	16.7%	75.0%	8.3%	12
<b>Measurement and Geometry</b>								
All students	63.0%	37.0%	—	135	12.7%	81.8%	5.5%	110
Females	70.8%	29.2%	—	72	11.3%	82.3%	6.5%	62
Males	52.9%	47.1%	—	51	12.9%	80.6%	6.5%	31
African Americans	65.0%	35.0%	—	100	13.2%	80.3%	6.6%	76
Whites	57.1%	42.9%	—	14	8.3%	83.3%	8.3%	12
<b>Data Analysis and Probability</b>								
All students	11.9%	75.6%	12.6%	135	50.0%	50.0%	—	110
Females	9.7%	79.2%	11.1%	72	48.4%	51.6%	—	62
Males	15.7%	70.6%	13.7%	51	54.8%	45.2%	—	31
African Americans	14.0%	73.0%	13.0%	100	50.0%	50.0%	—	76
Whites	7.1%	78.6%	14.3%	14	50.0%	50.0%	—	12

\* total number of students in Level 1

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

**TABLE 6.14**  
**Fall 2006 HSAP ELA Operational Test:**  
**Content-Area Information (All Attempts)**

<b>Subgroup</b>	<b>Level 1</b>				<b>Level 2 and Above</b>			
	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<b>N1*</b>	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	<b>N2**</b>
<b>Reading Process and Comprehension</b>								
All students	71.4%	26.8%	1.8%	4,004	6.0%	40.8%	53.2%	5,301
Females	68.7%	29.6%	1.6%	1,414	4.9%	40.0%	55.2%	2,445
Males	72.8%	25.3%	1.9%	2,519	7.2%	42.2%	50.6%	2,707
African Americans	72.5%	26.3%	1.1%	2,652	8.7%	54.7%	36.6%	2,392
Whites	67.8%	29.4%	2.8%	870	3.9%	27.9%	68.2%	2,261
<b>Analysis of Texts</b>								
All students	56.7%	38.0%	5.3%	4,004	3.3%	36.0%	60.7%	5,301
Females	54.8%	40.0%	5.2%	1,414	3.2%	37.2%	59.6%	2,445
Males	57.4%	37.1%	5.5%	2,519	3.5%	35.6%	60.8%	2,707
African Americans	55.3%	39.9%	4.8%	2,652	4.4%	48.3%	47.2%	2,392
Whites	57.5%	35.5%	7.0%	870	2.2%	24.4%	73.4%	2,261
<b>Word Study and Analysis</b>								
All students	53.5%	44.7%	1.8%	4,004	9.7%	70.7%	19.6%	5,301
Females	52.8%	45.8%	1.5%	1,414	11.3%	69.9%	18.8%	2,445
Males	53.9%	44.1%	2.0%	2,519	8.5%	71.6%	19.9%	2,707
African American	52.6%	45.9%	1.5%	2,652	13.0%	75.1%	11.8%	2,392
White	50.5%	46.2%	3.3%	870	5.6%	66.6%	27.8%	2,261
<b>Writing</b>								
All students	56.0%	36.3%	7.6%	4,004	3.1%	23.5%	73.4%	5,301
Females	48.2%	42.8%	9.1%	1,414	1.7%	21.0%	77.3%	2,445
Males	60.2%	32.9%	6.9%	2,519	4.3%	26.3%	69.4%	2,707
African Americans	54.1%	38.2%	7.7%	2,652	4.0%	31.9%	64.1%	2,392
Whites	56.3%	34.6%	9.1%	870	2.0%	16.2%	81.8%	2,261
<b>Research</b>								
All students	30.3%	66.8%	2.9%	4,004	8.0%	60.9%	31.1%	5,301
Females	31.5%	66.0%	2.5%	1,414	8.7%	61.4%	29.9%	2,445
Males	29.5%	67.2%	3.3%	2,519	7.5%	60.9%	31.6%	2,707
African Americans	32.2%	65.4%	2.4%	2,652	11.3%	71.4%	17.2%	2,392
Whites	27.4%	67.8%	4.8%	870	5.2%	50.9%	44.0%	2,261

\* total number of students in Level 1

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

**TABLE 6.15**  
**Spring 2007 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>	N1*	<i>Needs Improvement</i>	<i>May Need Improvement</i>	<i>Adequate</i>	N2**
<b>Reading Process and Comprehension</b>								
All students	51.8%	44.5%	3.7%	6,313	0.7%	16.5%	82.9%	45,762
Females	45.0%	50.8%	4.2%	2,200	0.5%	14.8%	84.6%	23,899
Males	55.4%	41.2%	3.5%	4,070	0.8%	18.2%	80.9%	21,742
African Americans	50.9%	45.4%	3.7%	3,795	1.0%	24.2%	74.9%	16,612
Whites	53.3%	42.9%	3.7%	1,875	0.5%	11.6%	87.9%	26,544
<b>Analysis of Texts</b>								
All students	58.8%	36.8%	4.5%	6,313	1.0%	16.7%	82.3%	45,762
Females	55.9%	40.1%	4.0%	2,200	0.9%	15.9%	83.2%	23,899
Males	60.2%	35.1%	4.6%	4,070	1.0%	17.6%	81.4%	21,742
African Americans	57.2%	38.4%	4.4%	3,795	1.1%	23.1%	75.8%	16,612
Whites	61.1%	34.6%	4.3%	1,875	0.8%	12.6%	86.6%	26,544
<b>Word Study and Analysis</b>								
All students	51.7%	45.7%	2.6%	6,313	1.9%	39.5%	58.6%	45,762
Females	50.5%	47.7%	1.8%	2,200	1.8%	40.5%	57.7%	23,899
Males	52.3%	44.6%	3.0%	4,070	2.0%	38.2%	59.8%	21,742
African Americans	53.3%	44.7%	1.9%	3,795	3.4%	55.2%	41.4%	16,612
Whites	46.0%	49.8%	4.2%	1,875	0.8%	29.1%	70.1%	26,544
<b>Writing</b>								
All students	63.9%	30.8%	5.3%	6,313	1.1%	12.6%	86.2%	45,762
Females	56.7%	36.9%	6.4%	2,200	0.7%	10.7%	88.7%	23,899
Males	67.8%	27.5%	4.7%	4,070	1.7%	14.7%	83.6%	21,742
African Americans	62.7%	32.2%	5.1%	3,795	1.7%	19.2%	79.1%	16,612
Whites	64.3%	29.6%	6.1%	1,875	0.7%	8.3%	91.0%	26,544
<b>Research</b>								
All students	40.0%	56.9%	3.1%	6,313	3.6%	46.1%	50.3%	45,762
Females	40.9%	57.0%	2.1%	2,200	4.0%	47.6%	48.4%	23,899
Males	39.6%	56.9%	3.6%	4,070	3.3%	44.4%	52.4%	21,742
African Americans	40.4%	57.0%	2.6%	3,795	5.9%	61.5%	32.6%	16,612
Whites	39.3%	56.2%	4.5%	1,875	2.2%	36.1%	61.7%	26,544

\* total number of students in Level 1

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

**TABLE 6.16**  
**Summer 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**
<b>Reading Process and Comprehension</b>								
All students	39.5%	48.8%	11.6%	43	43.8%	56.3%	—	16
Females	42.3%	46.2%	11.5%	26	37.5%	62.5%	—	8
Males	37.5%	50.0%	12.5%	16	50.0%	50.0%	—	8
African Americans	41.4%	44.8%	13.8%	29	36.4%	63.6%	—	11
Whites	25.0%	50.0%	25.0%	4	100.0%	—	—	1
<b>Analysis of Texts</b>								
All students	41.9%	53.5%	4.7%	43	50.0%	50.0%	—	16
Females	42.3%	50.0%	7.7%	26	37.5%	62.5%	—	8
Males	37.5%	62.5%	—	16	62.5%	37.5%	—	8
African Americans	44.8%	51.7%	3.4%	29	54.5%	45.5%	—	11
Whites	100.0%	—	—	4	100.0%	—	—	1
<b>Word Study and Analysis</b>								
All students	53.5%	46.5%	—	43	18.8%	62.5%	18.8%	16
Females	57.7%	42.3%	—	26	12.5%	75.0%	12.5%	8
Males	43.8%	56.3%	—	16	25.0%	50.0%	25.0%	8
African American	62.1%	37.9%	—	29	9.1%	63.6%	27.3%	11
White	50.0%	50.0%	—	4	100.0%	—	—	1
<b>Writing</b>								
All students	30.2%	58.1%	11.6%	43	6.3%	56.3%	37.5%	16
Females	30.8%	57.7%	11.5%	26	12.5%	37.5%	50.0%	8
Males	25.0%	62.5%	12.5%	16	75.0%	25.0%	—	8
African Americans	31.0%	58.6%	10.3%	29	9.1%	63.6%	27.3%	11
Whites	25.0%	50.0%	25.0%	4	100.0%	—	—	1
<b>Research</b>								
All students	30.2%	67.4%	2.3%	43	75.0%	25.0%	—	16
Females	34.6%	61.5%	3.8%	26	87.5%	12.5%	—	8
Males	18.8%	81.3%	—	16	62.5%	37.5%	—	8
African Americans	34.5%	65.5%	—	29	81.8%	18.2%	—	11
Whites	100.0%	—	—	4	100.0%	—	—	1

\* 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 6.17 for students overall and by gender and selected ethnic group.

**TABLE 6.17**  
**HSAP Summary Statistics Overall and by Subgroups**

Subgroup	Mathematics			ELA		
	Scale Score			Scale Score		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.
<b>Fall 2006</b>						
All students	12,658	198.6	21.0	9,332	205.2	24.1
Females	5,969	199.2	19.7	3,868	208.9	23.7
Males	6,256	197.2	21.3	5,239	202.2	23.6
African Americans	7,441	193.2	14.6	5,059	198.2	18.2
Whites	3,921	206.2	25.5	3,137	216.2	27.1
<b>Spring 2007</b>						
All students	52,202	222.3	29.0	52,152	226.4	22.6
Females	26,137	222.7	27.4	26,113	229.4	21.4
Males	25,863	222.0	30.5	25,867	223.5	23.5
African Americans	20,480	209.8	24.0	20,446	217.4	20.6
Whites	28,434	231.5	28.7	28,444	233.6	21.3
<b>Summer 2007</b>						
All students	245	197.4	9.7	59	195.9	12.0
Females	134	197.6	10.2	34	195.5	11.0
Males	82	196.3	9.3	24	197.1	13.2
African Americans	176	197.2	10.3	40	195.2	12.5
Whites	26	196.8	9.0	5	199.2	10.5

## 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 2006	Spring 2007	Summer 2007
<b>Mathematics</b>			
Reliability	0.91	0.95	0.70
SEM	3.07	3.22	2.97
<b>ELA</b>			
Reliability	0.95	0.95	0.84
SEM	5.07	4.08	4.07

#### 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 2006	6.46	5.41	5.84	7.55
Mathematics, Spring 2007	7.00	5.69	6.22	7.53
Mathematics, Summer 2007	5.37	5.41	5.81	7.56
ELA, Fall 2006	10.69	5.32	6.07	7.23
ELA, Spring 2007	8.99	5.20	6.09	7.46
ELA, Summer 2007	5.99	5.20	6.13	7.52

### 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
<b>Fall 2006</b>		
Mathematics	85.9	91.7
ELA	87.0	89.9
<b>Spring 2007</b>		
Mathematics	91.2	89.1
ELA	93.0	87.9
<b>Summer 2007</b>		
Mathematics	75.2	98.5
ELA	81.1	98.6

## 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 the PEM selected for detecting DIF were the Mantel-Haenszel (MH) chi-square for dichotomous items (MC items) and Mantel's chi-square for polytomous items (CR and ER items). The PEM 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 ( $\Delta$ ) 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 ( $\Delta$ ) 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 ( $\Delta$ ) 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 2006 and Spring 2007 operational forms, Tables 8.1 and 8.2 report the numbers of items in the various DIF categories for mathematics and ELA.

**TABLE 8.1**  
**Fall 2006 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
<b>Mathematics</b>						
Multiple choice	Male	Female	62	61	1	0
Multiple choice	White	Black	62	61	1	0
Constructed response	Male	Female	3	3	0	0
Constructed response	White	Black	3	3	0	0
<b>ELA</b>						
Multiple choice	Male	Female	60	58	2	0
Multiple choice	White	Black	60	56	2	2
Constructed response	Male	Female	2	2	0	0
Constructed response	White	Black	2	2	0	0
Extended response	Male	Female	4	4	0	0
Extended response	White	Black	4	4	0	0

**TABLE 8.2**  
**Spring 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
<b>Mathematics</b>						
Multiple choice	Male	Female	62	55	7	0
Multiple choice	White	Black	62	55	7	0
Constructed response	Male	Female	3	3	0	0
Constructed response	White	Black	3	3	0	0
<b>ELA</b>						
Multiple choice	Male	Female	60	57	2	1
Multiple choice	White	Black	60	57	3	0
Constructed response	Male	Female	2	2	0	0
Constructed response	White	Black	2	2	0	0
Extended response	Male	Female	4	4	0	0
Extended response	White	Black	4	4	0	0

**TABLE 8.3**  
**Summer 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
<b>Mathematics</b>						
Multiple choice	Male	Female	62	56	4	2
Multiple choice	White	Black	62	60	1	1
Constructed response	Male	Female	3	3	0	0
Constructed response	White	Black	3	3	0	0
<b>ELA</b>						
Multiple choice	Male	Female	60	60	0	0
Multiple choice	White	Black	60	60	0	0
Constructed response	Male	Female	2	2	0	0
Constructed response	White	Black	2	2	0	0
Extended response	Male	Female	4	4	0	0
Extended response	White	Black	4	4	0	0

#### 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 8.3 and 8.4 report the correlation matrices among the reporting category scores.

**TABLE 8.4**  
**Fall 2006 HSAP Correlations among Reporting Categories**

Mathematics (N = 12,658)						ELA (N = 9,332)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
<b>NO</b>	1.00	0.67	0.61	0.57	0.63	<b>RC</b>	1.00	0.74	0.60	0.63	0.57
<b>AL</b>	—	1.00	0.64	0.59	0.62	<b>AT</b>	—	1.00	0.62	0.59	0.52
<b>MG</b>	—	—	1.00	0.57	0.65	<b>WS</b>	—	—	1.00	0.48	0.42
<b>DP</b>	—	—	—	1.00	0.56	<b>WR</b>	—	—	—	1.00	0.39
<b>IR</b>	—	—	—	—	1.00	<b>RS</b>	—	—	—	—	1.00

**TABLE 8.5**  
**Spring 2007 HSAP Correlations among Reporting Categories**

Mathematics (N = 58,711)						ELA (N =55,690)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
NO	1.00	0.77	0.76	0.74	0.74	RC	1.00	0.76	0.69	0.69	0.64
AL	—	1.00	0.76	0.73	0.75	AT	—	1.00	0.69	0.69	0.61
MG	—	—	1.00	0.75	0.77	WS	—	—	1.00	0.64	0.59
DP	—	—	—	1.00	0.72	WR	—	—	—	1.00	0.53
IR	—	—	—	—	1.00	RS	—	—	—	—	1.00

**TABLE 8.6**  
**Summer 2007 HSAP Correlations among Reporting Categories**

Mathematics (N = 245)						ELA (N = 59)					
Reporting Category	NO	AL	MG	DP	IR	Reporting Category	RC	AT	WS	WR	RS
NO	1.00	0.40	0.25	0.36	0.24	RC	1.00	0.46	0.36	0.31	0.41
AL	—	1.00	0.18	0.34	0.19	AT	—	1.00	0.47	0.50	0.56
MG	—	—	1.00	0.30	0.30	WS	—	—	1.00	0.31	0.34
DP	—	—	—	1.00	0.27	WR	—	—	—	1.00	0.43
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. 2006a. *Fall 2006 HSAP Test Administration Manual*. Columbia: South Carolina Department of Education.

- . 2006b. *Fall 2006 HSAP District Test Coordinator's Supplement*. Columbia: South Carolina Department of Education.
- . 2007a. *Spring 2007 HSAP Test Administration Manual*. Columbia: South Carolina Department of Education.
- . 2007b. *Spring 2007 HSAP District Test Coordinator's Supplement*. Columbia: South Carolina Department of Education.
- . 2007c. *Summer 2007 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.