

# **The Connecticut Academic Performance Test: Technical Report**

**Prepared by  
Irene Hendrawan**

**September 2007**



# Table of Contents

<b>Table of Contents</b> .....	2
<b>List of Charts</b> .....	3
<b>List of Tables</b> .....	4
<b>Part 1: Introduction</b> .....	5
1.1. General description of CAPT.....	5
1.2. Test Design CAPT3.....	5
1.3. Test Forms CAPT3.....	6
<b>Part 2: Test Development</b> .....	7
<b>Part 3: Item Level Statistics</b> .....	8
<b>Part 4: Scaling and Equating</b> .....	14
4.1. 2007 CAPT Linking Items.....	14
4.2. Calibration Process.....	15
<b>Part 5: Test Statistics</b> .....	22
5.1. Reliability.....	22
5.2. Classification Consistency and Accuracy.....	22
<b>Part 6: CAPT3 Standards</b> .....	24
<b>Part 7: Validity</b> .....	26
7.1. Content Validity Survey.....	26
7.2. Scoring Quality Assurance Procedures Undertaken during Development.....	26
7.3. Item Quality Analysis Undertaken During Development.....	26
7.4. Equating Design.....	28
<b>References</b> .....	29
<b>Appendix A: Raw, Theta, and Scale Scores</b> .....	30

## List of Charts

<b>Chart 1: Calibration Design for 2007 Mathematics and Science</b> .....	15
<b>Chart 2: Calibration Design for 2007 Reading</b> .....	16
<b>Chart 3: Calibration Design for 2007 Writing</b> .....	16

## List of Tables

<b>Table 1: 2007 CAPT Operational Test Design</b> .....	5
<b>Table 2: 2007 CAPT Test Forms</b> .....	6
<b>Table 3: Mathematics (Main Form) Item Analysis</b> .....	9
<b>Table 4: Science (Main Form) Item Analysis</b> .....	10
<b>Table 5: Reading for Information (Main Form) Item Analysis</b> .....	12
<b>Table 6: Editing and Revising (Main Form) Item Analysis</b> .....	13
<b>Table 7: Response to Literature and Interdisciplinary Writing (Main Form) Item Analysis</b> .....	13
<b>Table 8: 2007 Internal (Embedded) Linking Items</b> .....	14
<b>Table 9: CAPT3 (2007) Equating Constants</b> .....	17
<b>Table 10: Summary of Weighting for Reading and Writing</b> .....	17
<b>Table 11: Scaling Coefficients from Base Year (CAPT2)</b> .....	18
<b>Table 12: Raw, Theta, and Scale Scores for Mathematics Live</b> .....	18
<b>Table 13: Raw, Theta, and Scale Scores for Science Live</b> .....	19
<b>Table 14: Raw, Theta, and Scale Scores for Reading Live</b> .....	20
<b>Table 15: Raw, Theta, and Scale Scores for Writing Live</b> .....	21
<b>Table 16: CAPT Cronbach’s Alpha</b> .....	22
<b>Table 17: Classification Consistency</b> .....	22
<b>Table 18: Classification Accuracy</b> .....	23
<b>Table 19: False Negative Classification</b> .....	23
<b>Table 20: False Positive Classification</b> .....	23
<b>Table 21: CAPT3 (2007) Achievement Levels and Scale Score Ranges</b> .....	25

# Part 1: Introduction

## 1.1. General description of CAPT

The Connecticut Academic Performance Test (CAPT) was designed to measure student performance in high school. Students are tested in the areas of Mathematics, Science, Reading, and Writing.

The CAPT has measured achievement of Connecticut students since 1994, when it was first administered. The second generation of CAPT was introduced in 2001. The content structure of the first generation CAPT was used as the baseline in developing the second generation. For the most part, the educational outcomes tested in the first generation were carried over to the second generation. Changes were made in light of new trends in instruction, educational assessment, and the lessons learned over the years of the first generation. The third generation of CAPT was introduced in the spring of 2007.

## 1.2. Test Design CAPT3

The current generation of the CAPT continues to be comprised four content areas:

1. Mathematics  
Mathematics (MA) has thirty-two operational (OP) -- twenty-four grid-in (GR) response items and eight open-ended (OE) items scored on 0-3 scale.
2. Science  
Science (SC) has sixty-five OP items -- sixty multiple choice (MC) items and five OE items scored on 0-3 scale.
3. Reading  
Reading (RD) consists of two subtests:
  - Reading for Information  
Reading for Information (RI) has eighteen OP items -- twelve MC items and six OE items scored on 0-2 scale.
  - Response to Literature  
Response to Literature (RL) consists of an extended response (EX) item with a 2-12 score scale (sum of two rater scores on a 1-6 scale).
4. Writing  
Writing (WR) consists of three subtests:
  - Editing & Revising  
Editing & Revising (ER) has eighteen MC items.
  - Interdisciplinary Writing 1 & Interdisciplinary Writing 2  
Interdisciplinary Writing 1 (IW1) & Interdisciplinary Writing 2 (IW2) have an EX item with a 2-12 score scale (sum of two rater scores on a 1-6 scale).

**Table 1: 2007 CAPT Operational Test Design**

Content Area	Subject	Number of Items				Total Items	Raw Score
		MC	GR	OE	EX		
Mathematics	Mathematics		24	8		32	0 - 48
Science	Science	60		5		65	0 - 75
Reading	Reading for Information	12		6		18	0 - 24
	Response to Literature				1	1	2 - 12
Writing	Editing & Revising	18				18	0 - 18
	Interdisciplinary Writing 1				1	1	2 - 12
	Interdisciplinary Writing 2				1	1	2 - 12

### 1.3. CAPT 2007 Test Forms

In the 2007 administration, two main forms were available for administration: Form HS14, which is the live form taken by most of the students, and Form HS0, which was available for breach situations. Moreover, Form HS0 will be used as a breach form in subsequent years of the third generation. Although the two forms were pre-equated during test assembly, there was still a need to carry out a post equating procedure after the test administration in order to ensure the comparability of the two forms.

Post equating was accomplished by using common test sessions or subtests that connected the two forms. Two stratified samples of 2,000 students were selected to take two hybrid forms that consisted of part of Form HS14 (Live) and part of Form HS0 (Breach). Scores on the hybrid forms constituted the live scores for the members of the two samples. Table 2 shows how these hybrid forms were built.

**Table 2: 2007 CAPT Test Forms**

Form	Mathematics	Science	Reading	Writing	N
HS14 - Live	Session 1 HS14	Session 1 HS14	RL HS14	IW1 HS14	Most of the State (approximately 40,000)
	Session 2 HS14	Session 2 HS14	RI HS14	IW2 HS14	
				ER HS14	
Hybrid 1	Session 1 HS14	Session 1 HS14	RL HS14	IW1 HS14	2,000
	Session 2 HS0	Session 2 HS0	RI HS0	IW2 HS14	
				ER HS0	
Hybrid 2	Session 1 HS0	Session 1 HS0	RL HS0	IW1 HS0	2,000
	Session 2 HS14	Session 2 HS14	RI HS14	IW2 HS0	
				ER HS14	
HS0 - Breach	Session 1 HS0	Session 1 HS0	RL HS0	IW1 HS0	Hybrid or Breach cases
	Session 2 HS0	Session 2 HS0	RI HS0	IW2 HS0	
				ER HS0	

The stratification of the samples for the hybrid forms was based on District Reference Groups (DRG). DRG categories are based on social and economic factors. MI selects a stratified sample of schools, based on the DRGs to which each belongs.

Any student who breaches a test session or subtest (HS14 or HS0) was given the corresponding test session or subtest (HS14 or HS0).

## Part 2: Test Development

The process by which each form of the CAPT is developed is extensive, spanning a five- or six-year period and many stages. The development process is led and overseen by staff members in the Bureau of Student Assessment at the Connecticut State Department of Education (CSDE), but it also involves many other people who represent a wide variety of perspectives and areas of expertise. CSDE curriculum specialists and content experts play a critical role and work closely with the assessment staff throughout the process. In addition, a major testing company and other organizations and individuals with experience in educational assessment are involved at appropriate points in the development process.

Advisory committees of Connecticut educators are particularly important throughout the development of the CAPT. Content Advisory and Fairness Committees review each item to ensure the match between the content objectives and the items, and to ensure meaningful interpretability of test results. The Content Advisory Committees included content experts, regular and special education teachers, Connecticut State Department of Education curriculum and assessment content specialists. A separate advisory committee is established for each part of the CAPT: Mathematics, Science, Reading, and Writing. These advisory committee members are selected on the basis of their knowledge in educational content and processes. In addition, the Fairness Committee is responsible for determining whether items are appropriate and fair to all examinees. Educators are carefully selected for the advisory committees to be representative of school districts throughout Connecticut.

The test development process for CAPT3 began with content specialists and testing experts writing test specifications with the help of the CAPT content advisory committees. The starting point for this process was looking at the specifications and structure of the first generation CAPT, and examining what has been working and what needed improvement. The new curriculum frameworks adopted by the State of Connecticut were also used as a guide. Test items for the CAPT3 were carefully developed in accordance with the established test specifications and test blueprint. These items were carefully matched to the content standards in the Connecticut Curriculum Frameworks for Mathematics, Science, Reading, and Writing. Items that did not pass the scrutiny of either Content Advisory or Fairness Committees were eliminated from the pool of pilot items.

After committee reviews, field test forms were created and piloted on a representative sample, stratified by District Reference Group (DRG), consisting of approximately 2000 students per form. Pilot statistics such as the mean, point biserial, and Rasch difficulty were generated and reviewed by CSDE assessment content staff and psychometricians. In addition, for hand-scored constructed response items, the contractor staff provided qualitative summaries about whether students appeared to have sufficient contextual knowledge to be able to fully respond to the item. Flawed items were removed from the item pool, including those showing test item bias or inappropriate levels of difficulty. Based on the CAPT3 Blueprints, four Mathematics, four Science, four Reading, and four Writing test forms of equivalent difficulty were simultaneously constructed from the pool of items that met all the review criteria. Every effort was made to ensure that strand level difficulties were comparable and that the items reflected the appropriate range of content within the strands across the generation.

## Part 3: Item Level Statistics

Tables 3-7 present item analysis (item quality) data for Mathematics, Science, Reading and Writing. The following information is presented in each item analysis:

**Classical and IRT difficulties:** The proportion of students who answer each item correctly (PC) and the IRT Rasch Difficulty (see section 4.2.) are presented. The proportion correct or  $p$ -value (for multiple-choice items) or mean item score (for constructed response items) are used to identify items that exceed acceptable difficulty range; too difficult or too easy. Generally, items are selected so that the  $p$ -values are greater than 0.25. This practice helps in discriminating among students. In addition, it is thought that very hard items may not align to the curriculum frameworks very well.

**Item Discriminations:** The point biserial correlation or item-total correlation is presented as measures of item discrimination. Item discrimination measures the relationship between the item score and the total score. The higher the correlation is the better the item discriminates. Point biserials for keyed responses (RPB) are presented for multiple-choice items. Item-total correlations are presented for constructed response items.

**Distractor Frequencies:** The proportion of students who answered each option (A-D, 0-3, and 2-12) are presented for the multiple-choice items, open-ended and extended response, respectively. The percent of students at each score point is presented for extended response (2-12).

**Table 3: Mathematics (Main Form) Item Analysis**

**Grid-in Items**

PC = Proportion Correct

RPB = Point-Biserial correlation

**Open-ended Items**

Mean = Mean OE score

Corr = Item-total correlation

0 – 3 = Percent of students at each score point

Order	Rasch	PC/Mean	RPB/Mean
1	-0.3799	0.54	0.50
2	-0.3157	0.53	0.59
3	0.1066	0.46	0.64
4	1.8424	0.20	0.47
5	0.9623	0.32	0.55
6	-0.7019	0.60	0.59
7	0.8691	0.33	0.62
8	0.8125	0.34	0.52
9	-0.2066	0.51	0.58
10	1.6000	0.23	0.58
11	-0.5522	0.57	0.51
12	-0.5274	0.57	0.50
13	0.5066	0.39	0.56
14	0.3959	0.41	0.51
15	-0.6782	0.59	0.48
16	0.4089	0.41	0.62
17	-0.2024	0.51	0.60
18	0.4298	0.40	0.61
19	-0.8866	0.63	0.46
20	0.9282	0.32	0.43
21	-0.4329	0.55	0.50
22	1.1660	0.28	0.56
23	-0.2879	0.53	0.50
24	-0.2411	0.52	0.62
25	0.0119	1.45	0.60
26	0.9655	0.87	0.68
27	-1.3085	2.14	0.61
28	0.4446	1.12	0.71
29	-0.5039	1.72	0.66
30	0.0336	1.43	0.59
31	-1.0484	1.94	0.60
32	0.1798	1.30	0.67

**Table 4: Science (Main Form) Item Analysis**

**Multiple-choice Items**

PC = Proportion Correct

RPB = Point-biserial correlation for keyed answer

**Open-ended Items**

Mean = Mean OE score

Corr = Item-total correlation

0 – 3 = Percent of students at each score point

Order	Rasch	PC/Mean	RPB/Corr
1	-0.6368	0.73	0.47
2	-0.7112	0.74	0.38
3	-0.5395	0.71	0.46
4	0.5883	0.49	0.30
5	0.5667	0.49	0.50
6	-0.2542	0.66	0.53
7	1.0954	0.39	0.26
8	0.3641	0.54	0.40
9	0.5723	0.49	0.39
10	0.6178	0.48	0.43
11	0.9091	0.42	0.27
12	0.7519	0.46	0.35
13	1.1803	0.37	0.34
14	-0.0041	0.61	0.4
15	0.2229	0.57	0.36
16	-0.5677	0.72	0.46
17	-0.2248	0.65	0.45
18	-0.0882	0.63	0.37
19	1.2942	0.35	0.36
20	0.5524	0.50	0.34
21	0.3248	0.54	0.40
22	0.3267	0.54	0.16
23	-0.3964	0.69	0.33
24	0.7262	0.46	0.43
25	-0.5048	0.71	0.39
26	0.4604	0.52	0.39
27	-0.4305	0.69	0.43
28	-0.2816	0.66	0.49
29	-0.7281	0.74	0.50
30	-0.5111	0.71	0.32
31	0.3288	0.54	0.31
32	0.2193	0.57	0.35
33	-0.8447	0.76	0.38
34	0.1061	0.59	0.31
35	0.5509	0.50	0.16
36	-1.1746	0.81	0.44
37	2.0113	0.22	0.31

**Table 4: Science Item (Main Form) Analysis (Continued)**

<b>Order</b>	<b>Rasch</b>	<b>PC/Mean</b>	<b>RPB/Corr</b>
38	-0.6119	0.72	0.28
39	1.6294	0.29	0.32
40	-0.2458	0.66	0.47
41	-0.2637	0.66	0.22
42	1.6773	0.28	0.35
43	0.6920	0.47	0.38
44	-0.2104	0.65	0.41
45	-0.5523	0.71	0.39
46	0.7209	0.46	0.39
47	0.7506	0.46	0.41
48	0.5874	0.49	0.42
49	-0.0943	0.63	0.36
50	0.5580	0.50	0.44
51	0.1513	0.58	0.34
52	-0.3307	0.67	0.56
53	0.7525	0.45	0.26
54	0.5610	0.49	0.37
55	-0.2141	0.65	0.39
56	-0.3368	0.67	0.35
57	0.1998	0.57	0.48
58	-0.0207	0.61	0.49
59	-0.5744	0.72	0.49
60	-0.0877	0.63	0.44
61	-0.3081	2.08	0.49
62	0.4692	1.60	0.62
63	0.6511	1.41	0.37
64	0.0984	1.87	0.46
65	0.4114	1.60	0.58

**Table 5: Reading for Information (Main Form) Item Analysis**

**Multiple-choice Items**

PC = Proportion Correct

RPB = Point-biserial correlation for keyed answer

**Open-ended Items**

Mean = Mean OE score

Corr = Item-total correlation

0 – 2 = Percent of students at each score point

Order	Rasch	PC/Mean	RPB/Corr
1	-1.4443	0.79	0.23
2	-0.8129	0.70	0.41
3	-0.7342	0.69	0.23
4	-0.7544	0.69	0.26
5	-0.7558	0.69	0.40
6	-0.0018	0.56	0.31
7	-1.2833	0.77	0.36
8	-0.1870	0.59	0.43
9	0.7769	0.41	0.34
10	-0.6421	0.67	0.34
11	-1.5265	0.80	0.51
12	-1.6824	0.82	0.51
13	0.6671	0.91	0.53
14	0.6184	0.93	0.50
15	0.7861	0.85	0.53
16	1.1143	0.74	0.55
17	1.1702	0.72	0.53
18	0.5204	0.93	0.58

**Table 6: Editing and Revising (Main Form) Item Analysis**

**Multiple-choice Items**

PC = Proportion Correct

RPB = Point-biserial correlation for keyed answer

Order	Rasch	PC	RPB
1	0.0558	0.70	0.45
2	-1.4331	0.89	0.37
3	0.1817	0.68	0.35
4	-2.1694	0.94	0.37
5	0.7604	0.57	0.43
6	0.6181	0.59	0.38
7	0.7168	0.58	0.19
8	-0.9479	0.84	0.27
9	1.7167	0.38	0.21
10	-0.0903	0.72	0.37
11	0.9715	0.52	0.24
12	-1.0287	0.85	0.40
13	-0.0350	0.71	0.41
14	1.9204	0.34	0.11
15	-1.3429	0.88	0.40
16	0.3708	0.64	0.42
17	-1.2482	0.87	0.41
18	1.2995	0.46	0.14

**Table 7: Response to Literature and Interdisciplinary Writing (Main Form) Item Analysis**

**Extended Response**

Mean = Mean EX score

Corr = Item-total correlation

2 – 12 = Percent of students at each point

	Rasch	Mean	Corr	2	3	4	5	6	7	8	9	10	11	12
RL	0.4440	7.05	0.65	0.02	0.01	0.05	0.06	0.21	0.20	0.28	0.11	0.05	0.00	0.00
IW1	1.0087	7.35	0.64	0.02	0.02	0.05	0.07	0.15	0.17	0.25	0.15	0.08	0.03	0.01
IW2	0.8757	7.56	0.62	0.01	0.01	0.04	0.06	0.14	0.16	0.27	0.16	0.11	0.03	0.01

## Part 4: Scaling and Equating

### 4.1 2007 CAPT Linking Items

The 2007 CAPT Mathematics, Science, and Editing & Revising tests were equated with the 2006 CAPT (HS13) subtests by embedding linking items which were carried over from the live 2006 (HS13) administration. Linking items were not counted toward students' scores.

The Live form of the 2007 CAPT (HS14) included:

- Mathematics – 2 forms of each session (A and B). Four grid items were added to each form.
- Science – 2 forms of each session (A and B). Four MC items were added to each form.
- Editing & Revising – 2 forms (A and B). 1 Passage and 6 MC items were added to each form.
- IWs – 1 form. Every IW has an EX item.
- RI – had an external (supplemental) form, SU71. This form included 3 passages with 4 MC and 2 OE items per passage. Any school that administered SU71 in 2006 should not have administered it again in 2007. Previous scoring materials for the OE items were used again in 2007.
- RL – 1 form. RL has an EX item.

The Breach form of the 2007 CAPT included:

- Mathematics – 1 form of each session (A). Four grid items were added.
- Science – 1 form of each session (A). Four MC items were added.
- Editing & Revising – 1 form (A). 1 Passage and 6 MC items were added.
- IWs – 1 form. Every IW has an EX item.
- RI – 1 form.
- RL – 1 form. RL has an EX item.

Table 8 indicates the linking items used as well as their positions on the 2006 and 2007 test forms.

**Table 8: 2007 Internal (Embedded) Linking Items**

Subject	Form	Cluster/Passage Name	2006		2007	
			Session	Position	Session	Position
Mathematics	A	Ms. Romero's Photograph	1	9	1	9
		Air Pressure	1	12	1	12
		Mathematicians	2	25	2	25
		Calculators Sold	2	28	2	28
	B	Ms. Rivera's Cabana	1	14	1	14
		700 Page Novel	1	15	1	15
		Mathematicians	2	26	2	27
		Food Stand Sales	2	35	2	35
Science	A	Genetics and Evolution	1	9	1	10
		Genetics and Evolution	1	11	1	11
		Energy Sources	1	14	2	43
		Earth's Resources	2	44	2	44
	B	Cells	1	19	1	20
		Cells	1	22	1	21
		Chemistry and Cars	2	46	2	46
		Chemistry and Cars	2	47	2	47
Editing & Revising	A	1 - Dogfight	1	1 - 6	1	1 - 6
	B	4 - Teacher of the Year	1	19 -24	1	19 -24

## 4.2. Calibration Process

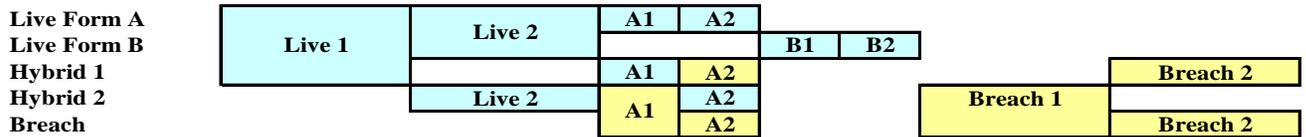
The CAPT 2007 test forms were scaled and equated using the Rasch model. The WINSTEPS software was used to estimate the latent trait difficulty of each item on the test. WINSTEPS, written by Linacre (Mesa Press, 2005) was used to complete Rasch analyses. WINSTEPS is a WINDOWS-based program that is widely used for similar high stakes tests. WINSTEPS (the Rasch model), allows for the estimation of item difficulty for multiple-choice, open-ended, and extended response items on a single scale. Using these item difficulties, the model is able to estimate the ability (theta) of each student and subsequently the student's raw score on the test.

All scaling and equating analyses were undertaken by three independent groups: Measurement Incorporated (MI), the contractor, the Connecticut State Department of Education (CSDE), and H. Jane Rogers and H. Swaminathan from the University of Connecticut (UCONN). Results were compared and cross-checked to the fourth decimal point to ensure accuracy.

The purpose of equating was to place the difficulty estimates of the items on the same scale as HS13; the last form of the second generation which ended in 2006. The equating was accomplished in the following steps:

1. For every content area, concurrently calibrate the 2007 OP with all forms (Live Form A, Live Form B, Hybrid 1, Hybrid 2, and Breach) combined (see Charts 1-3 for sample calibration data matrix). Thus, all forms are on the same scale. This step is a free run calibration. For RL, IW1, and IW2, 2 is subtracted from each score so that scores are on a scale from 0 to 10.

**Chart 1: Calibration Design for 2007 Mathematics and Science**



**Note:**

Live Form A = Live 1 + Live 2 + A1 + A2

Live Form B = Live 1 + Live 2 + B1 + B2

Hybrid 1 = Live 1 + A1 + A2 + Breach 2

Hybrid 2 = Live 2 + A1 + A2 + Breach 1

Breach = A1 + A2 + Breach 1 + Breach 2

Live 1 = HS14 Session 1

Live 2 = HS14 Session 2

A1 = Linking Items Form A in HS14/HS0 Session 1

A2 = Linking Items Form A in HS14/HS0 Session 2

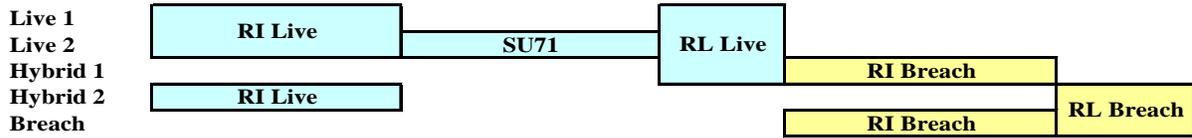
B1 = Linking Items Form B in HS14 Session 1

B2 = Linking Items Form B in HS14 Session 2

Breach 1 = HS0 Session 1

Breach 2 = HS0 Session 2

**Chart 2: Calibration Design for 2007 Reading**



**Note:**

- Live 1 = RI Live + RL Live
- Live 2 = RI Live + SU71 + RL Live
- Hybrid 1 = RL Live + RI Breach
- Hybrid 2 = RI Live + RL Breach
- Breach = RI Breach + RL Breach

- RI Live = RI HS14
- RL Live = RL HS14
- SU71 = External Linking Items (only 2,000 students will take this form)
- RI Breach = RI HS0
- RL Breach = RL HS0

**Chart 3: Calibration Design for 2007 Writing**



**Note:**

- Live Form A = ER Live + A + IW1 Live+ IW2 Live
- Live Form B = ER Live + B + IW1 Live + IW2 Live
- Hybrid 1 = A + IW1 Live + IW2 Live + ER Breach
- Hybrid 2 = ER Live + A + IW1 Breach + IW2 Breach
- Breach = A + ER Breach + IW1 Breach + IW2 Breach

- ER Live = ER HS14
- A = Linking Items Form A in ER Live/Breach
- B = Linking Items Form B in ER Live
- IW1 Live = IW1 HS14
- IW2 Live = IW2 HS14
- ER Breach = ER HS0
- IW1 Breach = IW1 HS0
- IW2 Breach = IW2 HS0

- Select the items linking HS14 (2007 live test form) and HS13 (2006 live form). Do anchor evaluation using .3 rule between the estimates of difficulties from Step 1 and HS13 values. This is an iterative process in which each item, starting with the one with the greatest absolute value difference, is removed until all items fulfill the criterion for inclusion. Using the remaining items the difference between the scale means from Step 1 and HS13 yields the equating constant. Table 9 shows the equating constants.

**Table 9: CAPT3 (2007) Equating Constants**

Content Area	Equating Constant
Mathematics	0.0566
Science	0.1552
Reading	-0.0331
Writing	0.0998

- Using the item output files from step 1 and anchoring these b-values, perform another run for each combination of forms, i.e., employ only those items from a given form in order to obtain theta values for each group of students administered a particular form. For Reading and Writing, the appropriate weights were included (see Table 10). Since the linking items do not contribute to the total test score, linking items (e.g., form A, form B, SU71) were deleted before generating the thetas.

**Table 10: Summary of Weighting for Reading and Writing**

Content/Subject	Unweighted Scale	% of Total Scale	Score Weight	Compute Formula	Weighted Scale
Reading for Information	0 - 24	50%	1.0		0 - 24
Response to Literature	2 - 12	50%	2.4	(RL - 2)*2.4	0 - 24
Total Reading	0 - 36				0 - 48
Editing & Revising	0 - 18	30%	1.0		0 - 18
Interdisciplinary Writing 1	2 - 12	35%	2.1	(IW1 - 2)*2.1	0 - 21
Interdisciplinary Writing 2	2 - 12	35%	2.1	(IW2 - 2)*2.1	0 - 21
Total Writing	0 - 42				0 - 60

- Compute scale score (SS) and scale score standard error (SSE) for each forms:

$$SS = \left( \frac{T + EQ - T_{mean}}{T_{SD}} \right) * 45 + 250 \text{ and } SSE = \frac{T_{err}}{T_{SD}} * 45$$

where

$T$  and  $T_{err}$  are the ability score and the standard error of the ability from the score file in Step 3.

$EQ$  is the difference between the mean of difficulty estimates of the linking items on HS13 and mean of difficulty estimates of the linking items on HS14, called the equating constant. This value was obtained in Step 2.

$T_{mean}$  and  $T_{SD}$  are the scaling coefficients from base year of CAPT2 (see Table 11).

**Table 11: Scaling Coefficients from Base Year (CAPT2)**

<b>Content Area</b>	<b>T_mean</b>	<b>T_SD</b>
Mathematics	-0.2317	1.6051
Science	0.4077	0.9254
Reading	0.4843	1.2278
Writing	1.0931	1.1187

The minimum SS is set to 100 and the maximum SS is set to 400. Any SS less than 100 was reset to 100 and any SS greater than 400 was reset to 400.

Tables 12-15 and Appendix A contain the results of raw scores, theta, and scale scores for all forms. Note that Combo1 is replacement form HS14 ER, HS14 IW1, and HS0 IW2, Combo2 is replacement form HS14 ER, HS0 IW1, and HS14 IW2, Combo3 is replacement form HS0 ER, HS14 IW1, and HS0 IW2, and Combo4 is replacement form HS0 ER, HS0 IW1, and HS14 IW2.

**Table 12: Raw, Theta, and Scale Scores for Mathematics Live**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.2169	112	25	0.0641	260
1	-4.0006	146	26	0.1545	262
2	-3.2874	166	27	0.2456	265
3	-2.8593	178	28	0.3373	268
4	-2.5470	187	29	0.4302	270
5	-2.2974	194	30	0.5242	273
6	-2.0871	200	31	0.6199	275
7	-1.9036	205	32	0.7174	278
8	-1.7395	209	33	0.8172	281
9	-1.5901	214	34	0.9199	284
10	-1.4522	217	35	1.0259	287
11	-1.3234	221	36	1.1362	290
12	-1.2023	224	37	1.2519	293
13	-1.0873	228	38	1.3743	297
14	-0.9776	231	39	1.5054	300
15	-0.8722	234	40	1.6478	304
16	-0.7705	236	41	1.8052	309
17	-0.6717	239	42	1.9827	314
18	-0.5755	242	43	2.1886	319
19	-0.4813	245	44	2.4360	326
20	-0.3886	247	45	2.7494	335
21	-0.2971	250	46	3.1827	347
22	-0.2063	252	47	3.9065	368
23	-0.1161	255	48	5.1339	400
24	-0.0260	257			

**Table 13: Raw, Theta, and Scale Scores for Science Live**

Raw Score	Theta	Scale Score	Raw Score	Theta	Scale Score
0	-5.5828	100	38	0.0499	240
1	-4.3680	100	39	0.1063	243
2	-3.6594	100	40	0.1629	246
3	-3.2384	100	41	0.2196	248
4	-2.9352	100	42	0.2767	251
5	-2.6963	107	43	0.3340	254
6	-2.4981	116	44	0.3918	257
7	-2.3279	125	45	0.4502	260
8	-2.1782	132	46	0.5092	262
9	-2.0439	138	47	0.5690	265
10	-1.9218	144	48	0.6297	268
11	-1.8094	150	49	0.6915	271
12	-1.7051	155	50	0.7545	274
13	-1.6074	160	51	0.8188	278
14	-1.5153	164	52	0.8847	281
15	-1.4279	168	53	0.9524	284
16	-1.3446	172	54	1.0220	287
17	-1.2648	176	55	1.0939	291
18	-1.1880	180	56	1.1681	295
19	-1.1139	184	57	1.2452	298
20	-1.0423	187	58	1.3256	302
21	-0.9726	190	59	1.4094	306
22	-0.9048	194	60	1.4972	311
23	-0.8386	197	61	1.5897	315
24	-0.7739	200	62	1.6874	320
25	-0.7104	203	63	1.7912	325
26	-0.6480	206	64	1.9022	330
27	-0.5866	209	65	2.0217	336
28	-0.5262	212	66	2.1515	342
29	-0.4665	215	67	2.2939	349
30	-0.4075	218	68	2.4523	357
31	-0.3491	221	69	2.6315	366
32	-0.2912	224	70	2.8392	376
33	-0.2337	226	71	3.0880	388
34	-0.1766	229	72	3.4016	400
35	-0.1198	232	73	3.8332	400
36	-0.0631	235	74	4.5528	400
37	-0.0066	237	75	5.7753	400

**Table 14: Raw, Theta, and Scale Scores for Reading Live**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.4902	100	25	0.1475	236
1	-4.3566	100	26	0.3041	242
2	-3.7241	100	27	0.4654	248
3	-3.3505	108	28	0.6324	254
4	-3.0752	118	29	0.8057	261
5	-2.8502	127	30	0.9861	267
6	-2.6545	134	31	1.1737	274
7	-2.4774	140	32	1.3686	281
8	-2.3129	146	33	1.5700	289
9	-2.1570	152	34	1.7772	296
10	-2.0071	157	35	1.9898	304
11	-1.8616	163	36	2.2079	312
12	-1.7188	168	37	2.4322	320
13	-1.5781	173	38	2.6640	329
14	-1.4382	178	39	2.9041	337
15	-1.2987	183	40	3.1524	347
16	-1.1590	189	41	3.4072	356
17	-1.0188	194	42	3.6659	365
18	-0.8778	199	43	3.9287	375
19	-0.7359	204	44	4.2023	385
20	-0.5928	209	45	4.5045	396
21	-0.4486	215	46	4.8774	400
22	-0.3028	220	47	5.4611	400
23	-0.1553	225	48	6.4833	400
24	-0.0054	231			

**Table 15: Raw, Theta, and Scale Scores for Writing Live**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.1022	100	31	0.4395	228
1	-3.8834	100	32	0.5436	232
2	-3.1770	100	33	0.6500	236
3	-2.7643	100	34	0.7589	241
4	-2.4722	111	35	0.8706	245
5	-2.2457	120	36	0.9852	250
6	-2.0598	127	37	1.1030	254
7	-1.9011	134	38	1.2239	259
8	-1.7615	139	39	1.3480	264
9	-1.6358	144	40	1.4752	269
10	-1.5206	149	41	1.6052	275
11	-1.4133	153	42	1.7379	280
12	-1.3120	157	43	1.8730	285
13	-1.2152	161	44	2.0103	291
14	-1.1220	165	45	2.1497	297
15	-1.0312	169	46	2.2914	302
16	-0.9423	172	47	2.4356	308
17	-0.8546	176	48	2.5827	314
18	-0.7675	179	49	2.7335	320
19	-0.6806	183	50	2.8889	326
20	-0.5935	186	51	3.0506	333
21	-0.5059	190	52	3.2202	340
22	-0.4175	193	53	3.4007	347
23	-0.3280	197	54	3.5960	355
24	-0.2375	200	55	3.8126	363
25	-0.1455	204	56	4.0617	373
26	-0.0521	208	57	4.3649	386
27	0.0428	212	58	4.7718	400
28	0.1393	216	59	5.4464	400
29	0.2375	220	60	6.6189	400
30	0.3375	224			

## Part 5: Test Statistics

### 5.1. Reliability

Table 16 summarizes reliability estimates for CAPT Mathematics, Science, Reading, and Writing. The reliability coefficients are based on Cronbach's alpha measure of internal consistency. When evaluating these results it is important to remember that reliability is partially a function of test length and thus reliability is likely to be greater for clusters that have more items. Within each content area the reliability estimates across the forms were very similar.

**Table 16: CAPT Cronbach's Alpha**

Content Area	Form	Cronbach's Alpha
Mathematics	Live Form A	0.9438
	Live Form B	0.9444
	Hybrid 1	0.9500
	Hybrid 2	0.9290
Science	Live Form A	0.9285
	Live Form B	0.9343
	Hybrid 1	0.9239
	Hybrid 2	0.9204
Reading	Live 1	0.8429
	Live 2	0.8217
	Hybrid 1	0.8690
	Hybrid 2	0.8461
Writing	Live Form A	0.8265
	Live Form B	0.8101
	Hybrid 1	0.8137
	Hybrid 2	0.8176

### 5.2. Classification Consistency and Accuracy

Classification consistency (see Table 17) and accuracy (see Table 18) were measured using the IRT-Class program developed by [CASMA](#) (Center for Advanced Studies in Measurement and Assessment) at the University of Iowa. The decision consistency and accuracy was assessed based on the given ability distribution and the difficulty of the items (IRT parameters).

**Table 17: Classification Consistency**

Content	Overall Classification Consistency	Cut 1	Cut 2	Cut 3	Cut 4
Mathematics	0.72090	0.93976	0.93014	0.92077	0.92476
Reading	0.88054	0.91461	0.93623	0.93573	0.93625
Science	0.68224	0.93906	0.92088	0.89789	0.91081
Writing	0.87304	0.93818	0.93831	0.93463	0.93610

**Table 18: Classification Accuracy**

<b>Content</b>	<b>Overall Classification Accuracy</b>	<b>Cut 1</b>	<b>Cut 2</b>	<b>Cut 3</b>	<b>Cut 4</b>
Mathematics	0.79419	0.95565	0.94805	0.94408	0.94582
Reading	0.85916	0.92633	0.94179	0.95510	0.94875
Science	0.76511	0.95714	0.94353	0.92669	0.93564
Writing	0.88296	0.95469	0.95475	0.95281	0.95324

The results of the program show that for the most part, classifications are highly consistent (see Table 17). The consistency ratings at each cut score are generally in the upper 90s. This tends to tail off at the highest cut score (i.e., the upper end of the distributions). The cumulative effect of applying all cut scores simultaneously yields an average consistency of around low to mid 80s. The classification accuracy examinations show (see Table 18), similarly, that the accuracy ratings at each cut score are generally in the upper 90s.

The program also computes the false negative rates for the test, which in effect are an estimate of those students that may have been misclassified in a performance category lower than their true performance category. The results of the false negatives, found in Table 19, indicate that a very small number of students may have been negatively misclassified in this way. Table 20 shows the false positive classification.

**Table 19: False Negative Classification**

<b>Content</b>	<b>Overall False Negative</b>	<b>Cut 1</b>	<b>Cut 2</b>	<b>Cut 3</b>	<b>Cut 4</b>
Mathematics	0.07778	0.01592	0.01677	0.02642	0.01885
Reading	0.08370	0.05787	0.05108	0.01955	0.01013
Science	0.09733	0.02222	0.02617	0.02835	0.02137
Writing	0.06769	0.03105	0.01228	0.03266	0.03170

**Table 20: False Positive Classification**

<b>Content</b>	<b>Overall False Positive</b>	<b>Cut 1</b>	<b>Cut 2</b>	<b>Cut 3</b>	<b>Cut 4</b>
Mathematics	0.12803	0.02843	0.03518	0.02950	0.03533
Reading	0.05714	0.01580	0.00713	0.02535	0.04112
Science	0.13756	0.02064	0.03030	0.04496	0.04298
Writing	0.04935	0.01426	0.03297	0.01453	0.01506

## Part 6: CAPT3 Standards

When standards were being established for first generation CAPT, a judgmental standard setting process called Modified Angoff (1971) was employed. Through that process, groups of educators who were familiar with the performance of students at a particular grade level in a particular content area were asked to predict how students who just meet a particular standard (e.g., goal standard) would perform on many different CAPT items. Using the judgment of these groups of educators in consideration with other validity checks, appropriate state goal and remedial standards were recommended by the Department and adopted by the State Board of Education. For the second generation CAPT (CAPT2), the standards were set using a method called Book Mark. In the procedure, all items in the test are arranged from easiest to most difficult. Then a group of educators are asked to mark up to the item at which a student at specific standard could respond to correctly. As in the first generation, the standards set by using the Book Mark method were adopted by the State Board of Education.

The third generation (CAPT3) standards were developed by carrying over the CAPT2 standards as well as department staff working with a CAPT3 Standards Advisory Panel composed of technical experts, district content experts and district research and testing specialists. The CAPT3 standards were set to be as rigorous as the CAPT2 standards and to be equivalent across grade levels and across content areas as much as possible. Transferring the standards allowed the Department to maintain the same performance standards for NCLB purposes. The purpose of this section is to summarize the procedures used to accomplish the task of carrying over the standards (see Cizek and Bunch, 2007, for a discussion of standard setting procedures). In all content areas, the standards define the different academic performance levels. The state goal has been an important benchmark for judging the quality of education in Connecticut for more than a decade. The proficient standard is used for accountability purposes as required by No Child Left Behind (NCLB) to make determinations about Adequate Yearly Progress (AYP) and schools in need of improvement.

To continue to comply with the NCLB accountability requirements, the Connecticut State Department of Education (CSDE) carried over from the CAPT2 to the CAPT3 the following previously adopted achievement standards: Below Basic, Basic, Proficient, Goal and Advanced. The process of carrying over on the standards was accomplished with an intergeneration linking study which included the equating of CAPT2 forms and CAPT3 forms. In addition to statistically linking the test generations, historical results from past CAPT2 administrations were taken into consideration as well as input from the CAPT Standards Review Panel composed of a diverse group of Connecticut educators, including curriculum directors, teachers and administrators.

The Standards Review Panel assisted in the identification of acceptable and valid test standards for each content area of CAPT3. The CAPT Standards Review Panel was given an overview of the CAPT3 including the content covered, score weighting, and reporting conventions. Differences between CAPT2 and CAPT3 were also discussed. Copies of the complete CAPT3 test booklets were available for reference. In addition, the procedures for carrying CAPT2 standards over to CAPT3 were presented in detail so that committee members would better understand their role in the process. They reviewed data from several related analyses and discussed implications from both an educational perspective and a technical perspective. They were asked particularly to provide input in the following three areas:

- Review the content of the CAPT, score weighting, and reporting conventions.
- Review results from the inter-generational linking procedure to ensure that standards are reasonable and appropriate across content area; and
- Provide subjective input about the reasonableness and consistency of the standards for all content areas based on their content expertise and historical results from past test administrations.

All procedures were discussed with and approved by the Technical Advisory Committee (TAC) prior to implementation. The TAC is composed of nationally recognized experts in the measurement field. Finally,

standards proposed by the standards review panel were presented to the State Board of Education for final approval. Standards were established based on scale scores (100-400) in four content areas: Mathematics, Science, Reading, and Writing.

Table 21 shows the range of scale scores in each performance category.

**Table 21: CAPT3 (2007) Achievement Levels and Scale Score Ranges**

Content Area	Scale Score Ranges				
	Below Basic	Basic	Proficient	Goal	Advanced
Mathematics	100 - 193	194 - 220	221 - 259	260 - 289	290 - 400
Science	100 - 189	190 - 214	215 - 264	265 - 294	295 - 400
Reading	100 - 173	174 - 204	205 - 250	251 - 282	283 - 400
Writing	100 - 182	183 - 211	212 - 249	250 - 290	291 - 400

## **Part 7: Validity**

According to the 1999 AERA, APA, NCME *Standards*, “It is helpful to consider the four phases leading from the original statement of purpose(s) to the final product: (a) delineation of the purpose(s) of the test and the scope of the construct or the extent of the domain to be measured; (b) development and evaluation of the test specifications; (c) development, field testing, evaluation, and selection of the items and scoring guides and procedures; and (d) the assembly and evaluation of the test for operational use.

In the development and maintenance of CAPT each of these phases is carefully planned and implemented. The following section details the critical psychometric procedures undertaken to ensure a strong validity argument for the use and interpretation of CAPT (Kane, 2006; Messick, 1989).

### **7.1. Content Validity Survey**

In order for the CAPT to serve its intended purposes, it is critical that users of the test results be confident that those results are meaningful. The test must measure those competencies that are critical to the decisions the test scores are informing.

A content validation study was conducted to examine the content validity of the CAPT for its intended applications. For this study, a survey of the strands proposed for the second generation CAPT was sent to approximately 4,000 Connecticut educators, parents, and other citizens. The purpose of the survey was to determine 1) the importance of the proposed Mathematics, Science, Reading Across the Disciplines, and Writing Across the Disciplines strands and 2) whether the strands are taught prior to the end of the 10<sup>th</sup> grade. The respondents characterized the strands as important educational outcomes to which students would be instructed prior to testing.

### **7.2. Scoring Quality Assurance Procedures Undertaken during Development**

Much of the following discussion applies to procedures undertaken during field testing and test construction phases of development work. Of course quality control is applied during the operational administration, but not with the aim of selecting or removing items.

In order to ensure the validity of inferences made from the CAPT tests there are quality control procedures in place for the scoring of the test. One such quality assurance component is to check the MC answer keys for MC items several times prior to test administration and one final time during the first run of live results. Items yielding low point biserial correlations are checked a final time for miskeying.

For constructed-response (CR) items, CAPT staff and contractor staff work with Connecticut educators to establish score boundaries in a process known as “range finding”. The score point examples and training sets so established are carried forward into operational scoring and elaborated with new samples of student responses. Reader training lasts up to several days, and readers must qualify by matching scores to several sets of prescored student responses. Once scoring begins, validity packets are used to maintain reader accuracy. These are packets of student responses with scores pre-assigned by CAPT staff and Connecticut educators. Readers periodically receive these packets, and their responses are compared to the pre-assigned scores. If a reader assigns too many discrepant scores, that reader is retrained or removed from the project. Other QA procedures include a 100% second read for the writing prompts (IW). There is a 20% second read for short answer and extended response items in mathematics and reading comprehension.

### **7.3. Item Quality Analysis Undertaken During Development**

Another part of assessing the quality and validity of inferences made from an instrument is to assess the quality of the items on the test. This quality is typically assessed by examining the classical item statistics as well as the potential for item bias. Item bias could lead to less valid inferences made for certain subgroups.

*Item specifications.* CAPT employs *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999) as a primary source of guidance in the construction, field testing, and documentation of the tests. The introduction to the 1999 *Standards* best describes how those *Standards* are and will be used in the development and evaluation of CAPT tests:

Evaluating the acceptability of a test or test application does not rest on the literal satisfaction of every standard in this document, and acceptability cannot be determined by using a checklist. (*Standards*, p. 4)

Thus, the terms ‘target’ and ‘goal’ are used when referring to various psychometric properties of the tests. For example, while it is a goal of test development for each high school test to have a reliability coefficient of .90 or greater, it is not our intention to scrap a test with a reliability coefficient of .89. Instead, the test results would be published, along with the reliability coefficient and associated standard error of measurement.

*Item statistics.* Because the CAPT tests are used in making individual decisions about students, they must be very reliable, particularly at cut points (the score points that separate adjacent achievement categories). Target reliability coefficients of .90 (or higher) are therefore set for the important cut points of each test.

Other psychometric properties include item difficulty, item discrimination, and differential item functioning. General statistical targets are provided below:

*For Multiple-Choice (MC) Items*

Percent correct: greater than or equal to .25  
Point biserial correlation with total score: greater than or equal to .20  
Mantel-Haenszel: No Category C items (see below)

*For Constructed-Response (CR) Items*

Difficulty: any level as long as all score points are well represented  
Correlation with total score: greater than or equal to .20  
Generalized Mantel-Haenszel: No chi-square significant at .05 level of alpha

It should be pointed out that the point biserial correlations for MC items and the correlations for CR items refer to total scores of the field test form with the influence of the item in question removed.

*Differential item functioning.* Items that disadvantage any identifiable subgroup of students are said to be biased and detract from the validity of the tests. While only human judges can determine whether or not an item is biased, item statistics can serve as a tool to help judges in their decisions. After field testing, a committee reviews item statistics that detect differential item functioning (DIF). Specifically, Mantel-Haenszel statistics are used as measures DIF.

Since its introduction in the field of epidemiology in 1959, Mantel-Haenszel statistics have been employed by many test developers, and several refinements have been added. Educational Testing Service (ETS) uses the Mantel-Haenszel statistic and calculates a D statistic which permits grouping of test items into three categories (Zieky, 1993). The D statistic is a function of the case-control odds estimator of risk generated by SAS’s PROC FREQ. The D statistic is calculated as follows:

1.  $\alpha$  = case-control estimate of risk (odds ratio)
2.  $\beta$  = natural log of  $\alpha$
3.  $D = -2.35 * \beta$

Camilli and Shepard (1994, p. 121) describe three categories of items with respect to D:

- A D does not significantly differ from zero using Mantel-Haenszel chi-square, or D’s absolute value is less than 1

- B D significantly differs from 0 and D has either (a) an absolute value less than 1.5 or (b) an absolute value not significantly different from 1
- C D's absolute value is significantly greater than or equal to 1.5

Camilli and Shepard note that Category B items are typically investigated for potential bias, while Category C items are typically removed. Others treat Category C items only as candidates for elimination, pending a reprieve from the committee. In other words, Category C items are considered unusable unless specifically declared usable by the committee. It should be noted that an item that allowed a target group to break out of a pattern of trailing behind the reference group on all other items would tend to fall into Category C. The committee would likely want to keep such an item, in spite of its Mantel-Haenszel status.

DIF occurs when an item shows different results by group (e.g., by race, socioeconomic class, or sex) that cannot be explained by known differences in the overall achievement levels of the two groups. Overall achievement level is typically taken as scores on an operational test, assuming that the operational test is itself free of bias. While committee members are free to examine all field-tested items, they must review all items with a Category C rating. Unless the committee specifically calls for the inclusion of any such item, that item is removed from the pool.

#### **7.4. Equating Design**

A different CAPT form is used each year. In order to ensure that appropriate comparisons can be made from one form of the CAPT to another, test forms must be equivalent to each other. Care must be taken when test items are developed, when items are selected to create forms, when tests are administered, and when tests are scored to keep all conditions as similar as possible for one test form to another. Two important characteristics that must be similar across forms are the content that is measured and the difficulty of the test.

Part 4 of this report details the procedures used to equate and scale the CAPT tests. As mentioned above, three independent groups undertake the analyses and cross-check all analyses and results to ensure accuracy. Connecticut expends great effort and resources to maintain an assessment program that employs high quality psychometric standards and quality assurance.

## REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: AERA.
- Angoff, W. H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (2<sup>nd</sup> ed., pp. 508-600), Washington, DC: American Council on Research.
- Camilli, G., & Shepard, L. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks, CA: Sage.
- Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4<sup>th</sup> ed., pp. 18-64). Westport, CT: American Council on Education/Praeger.
- Linacre, J. M., & Wright, B. D. (1993, 2006). *A user's guide to BIGSTEPS*. Chicago, IL: MESA Press.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22, 719-748.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed., pp. 13-103). New York: American Council on Education/Macmillan Publishing Company.
- Winsteps. (1991-2006©). Linacre, John M.
- Zieky, M. (1993). Practical questions in the use of DIF statistics in item development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning: Theory and practice* (pp. 337-364). Hillsdale, NJ: Lawrence Erlbaum Associates.

**Appendix A: Raw, Theta, and Scale Scores**

**Raw, Theta, and Scale Scores for Mathematics Hybrid 1**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.4674	105	25	0.0713	260
1	-4.1918	141	26	0.1617	263
2	-3.4178	162	27	0.2528	265
3	-2.9491	175	28	0.3448	268
4	-2.6091	185	29	0.4380	270
5	-2.3404	192	30	0.5325	273
6	-2.1165	199	31	0.6287	276
7	-1.9232	204	32	0.7268	278
8	-1.7519	209	33	0.8272	281
9	-1.5970	213	34	0.9302	284
10	-1.4550	217	35	1.0367	287
11	-1.3231	221	36	1.1472	290
12	-1.1996	224	37	1.2628	293
13	-1.0827	228	38	1.3850	297
14	-0.9716	231	39	1.5156	301
15	-0.8652	234	40	1.6572	305
16	-0.7627	237	41	1.8133	309
17	-0.6636	239	42	1.9892	314
18	-0.5671	242	43	2.1928	320
19	-0.4729	245	44	2.4373	326
20	-0.3802	247	45	2.7471	335
21	-0.2889	250	46	3.1760	347
22	-0.1984	253	47	3.8947	367
23	-0.1084	255	48	5.1181	400
24	-0.0186	258			

**Raw, Theta, and Scale Scores for Mathematics Hybrid 2**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.1776	113	25	0.0674	260
1	-3.9571	147	26	0.1537	262
2	-3.2383	167	27	0.2403	265
3	-2.8058	179	28	0.3277	267
4	-2.4902	188	29	0.4161	270
5	-2.2387	195	30	0.5059	272
6	-2.0278	201	31	0.5977	275
7	-1.8450	206	32	0.6919	277
8	-1.6825	211	33	0.7892	280
9	-1.5355	215	34	0.8901	283
10	-1.4006	219	35	0.9956	286
11	-1.2751	222	36	1.1066	289
12	-1.1574	226	37	1.2244	292
13	-1.0458	229	38	1.3504	296
14	-0.9394	232	39	1.4866	300
15	-0.8372	235	40	1.6356	304
16	-0.7384	237	41	1.8007	309
17	-0.6427	240	42	1.9872	314
18	-0.5492	243	43	2.2025	320
19	-0.4578	245	44	2.4596	327
20	-0.3680	248	45	2.7821	336
21	-0.2794	250	46	3.2235	348
22	-0.1919	253	47	3.9537	369
23	-0.1051	255	48	5.1841	400
24	-0.0187	258			

**Raw, Theta, and Scale Scores for Mathematics Breach**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.4363	106	25	0.0743	260
1	-4.1539	142	26	0.1606	263
2	-3.3712	164	27	0.2472	265
3	-2.8957	177	28	0.3347	267
4	-2.5511	187	29	0.4235	270
5	-2.2797	194	30	0.5138	272
6	-2.0550	200	31	0.6062	275
7	-1.8623	206	32	0.7009	278
8	-1.6929	211	33	0.7988	280
9	-1.5408	215	34	0.9003	283
10	-1.4021	219	35	1.0064	286
11	-1.2738	222	36	1.1178	289
12	-1.1539	226	37	1.2358	293
13	-1.0408	229	38	1.3617	296
14	-0.9331	232	39	1.4975	300
15	-0.8300	235	40	1.6455	304
16	-0.7307	238	41	1.8093	309
17	-0.6346	240	42	1.9938	314
18	-0.5410	243	43	2.2067	320
19	-0.4495	245	44	2.4606	327
20	-0.3599	248	45	2.7795	336
21	-0.2716	250	46	3.2166	348
22	-0.1843	253	47	3.9421	369
23	-0.0978	255	48	5.1691	400
24	-0.0117	258			

**Raw, Theta, and Scale Scores for Science Hybrid 1**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.6652	100	38	0.0070	238
1	-4.4462	100	39	0.0623	241
2	-3.7317	100	40	0.1176	243
3	-3.3051	100	41	0.1730	246
4	-2.9966	100	42	0.2287	249
5	-2.7527	104	43	0.2848	252
6	-2.5499	114	44	0.3413	254
7	-2.3754	122	45	0.3982	257
8	-2.2217	130	46	0.4558	260
9	-2.0840	136	47	0.5141	263
10	-1.9587	142	48	0.5733	266
11	-1.8437	148	49	0.6334	269
12	-1.7370	153	50	0.6947	272
13	-1.6373	158	51	0.7573	275
14	-1.5436	163	52	0.8214	278
15	-1.4549	167	53	0.8871	281
16	-1.3707	171	54	0.9547	284
17	-1.2903	175	55	1.0244	288
18	-1.2132	179	56	1.0964	291
19	-1.1390	182	57	1.1713	295
20	-1.0674	186	58	1.2490	298
21	-0.9981	189	59	1.3303	302
22	-0.9307	192	60	1.4155	307
23	-0.8651	196	61	1.5051	311
24	-0.8011	199	62	1.5998	316
25	-0.7385	202	63	1.7006	320
26	-0.6771	205	64	1.8083	326
27	-0.6167	208	65	1.9245	331
28	-0.5573	211	66	2.0506	337
29	-0.4987	213	67	2.1894	344
30	-0.4408	216	68	2.3438	352
31	-0.3836	219	69	2.5189	360
32	-0.3268	222	70	2.7222	370
33	-0.2706	225	71	2.9664	382
34	-0.2146	227	72	3.2751	397
35	-0.159	230	73	3.7016	400
36	-0.1035	233	74	4.4157	400
37	-0.0482	235	75	5.6344	400

**Raw, Theta, and Scale Scores for Science Hybrid 2**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.7404	100	38	-0.0155	237
1	-4.5094	100	39	0.0407	240
2	-3.7816	100	40	0.0973	242
3	-3.3453	100	41	0.1542	245
4	-3.0297	100	42	0.2116	248
5	-2.7808	102	43	0.2697	251
6	-2.5744	113	44	0.3285	254
7	-2.3976	121	45	0.3881	257
8	-2.2423	129	46	0.4486	260
9	-2.1036	135	47	0.5102	263
10	-1.9779	142	48	0.5729	266
11	-1.8627	147	49	0.6369	269
12	-1.7561	152	50	0.7024	272
13	-1.6566	157	51	0.7694	275
14	-1.5632	162	52	0.8381	278
15	-1.4749	166	53	0.9087	282
16	-1.3911	170	54	0.9815	285
17	-1.3110	174	55	1.0565	289
18	-1.2343	178	56	1.1340	293
19	-1.1605	181	57	1.2144	297
20	-1.0892	185	58	1.2979	301
21	-1.0202	188	59	1.3850	305
22	-0.9531	191	60	1.4761	310
23	-0.8878	195	61	1.5718	314
24	-0.8241	198	62	1.6726	319
25	-0.7618	201	63	1.7795	324
26	-0.7006	204	64	1.8935	330
27	-0.6405	207	65	2.0159	336
28	-0.5813	209	66	2.1485	342
29	-0.5229	212	67	2.2937	349
30	-0.4652	215	68	2.4548	357
31	-0.4081	218	69	2.6366	366
32	-0.3514	221	70	2.8469	376
33	-0.2951	223	71	3.0981	388
34	-0.2390	226	72	3.4141	400
35	-0.1832	229	73	3.8480	400
36	-0.1273	232	74	4.5699	400
37	-0.0715	234	75	5.7941	400

**Raw, Theta, and Scale Scores for Science Breach**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.8120	100	38	-0.0566	235
1	-4.5798	100	39	-0.0018	238
2	-3.8494	100	40	0.0532	240
3	-3.4096	100	41	0.1087	243
4	-3.0901	100	42	0.1645	246
5	-2.8372	100	43	0.2210	248
6	-2.6267	110	44	0.2781	251
7	-2.4459	119	45	0.3360	254
8	-2.2869	127	46	0.3947	257
9	-2.1446	133	47	0.4545	260
10	-2.0158	140	48	0.5153	263
11	-1.8977	145	49	0.5774	266
12	-1.7886	151	50	0.6408	269
13	-1.6871	156	51	0.7057	272
14	-1.5918	160	52	0.7723	275
15	-1.5021	165	53	0.8408	279
16	-1.4171	169	54	0.9112	282
17	-1.3363	173	55	0.9839	286
18	-1.2590	176	56	1.0591	289
19	-1.1849	180	57	1.1370	293
20	-1.1136	184	58	1.2181	297
21	-1.0447	187	59	1.3026	301
22	-0.9779	190	60	1.3910	305
23	-0.9131	193	61	1.4839	310
24	-0.8500	196	62	1.5819	315
25	-0.7883	199	63	1.6859	320
26	-0.7279	202	64	1.7968	325
27	-0.6687	205	65	1.9160	331
28	-0.6104	208	66	2.0453	337
29	-0.5531	211	67	2.1870	344
30	-0.4964	214	68	2.3445	352
31	-0.4404	216	69	2.5227	360
32	-0.3848	219	70	2.7290	370
33	-0.3297	222	71	2.9761	382
34	-0.2749	224	72	3.2879	398
35	-0.2203	227	73	3.7175	400
36	-0.1657	230	74	4.4348	400
37	-0.1112	232	75	5.6557	400

**Raw, Theta, and Scale Scores for Reading Hybrid 1**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.4941	100	25	0.2217	239
1	-4.3538	100	26	0.3711	245
2	-3.7156	100	27	0.5234	250
3	-3.3379	109	28	0.6795	256
4	-3.0591	119	29	0.8402	262
5	-2.8304	127	30	1.0062	268
6	-2.6308	135	31	1.1782	274
7	-2.4496	141	32	1.3569	281
8	-2.2805	147	33	1.5424	288
9	-2.1198	153	34	1.7354	295
10	-1.9649	159	35	1.9362	302
11	-1.8139	165	36	2.1456	310
12	-1.6657	170	37	2.3646	318
13	-1.5191	175	38	2.5946	326
14	-1.3736	181	39	2.8367	335
15	-1.2285	186	40	3.0905	344
16	-1.0836	191	41	3.3534	354
17	-0.9388	197	42	3.6212	364
18	-0.7940	202	43	3.8925	374
19	-0.6494	207	44	4.1727	384
20	-0.5049	213	45	4.4793	395
21	-0.3606	218	46	4.8546	400
22	-0.2161	223	47	5.4376	400
23	-0.0712	228	48	6.4543	400
24	0.0745	234			

**Raw, Theta, and Scale Scores for Reading Hybrid 2**

Raw Score	Theta	Scale Score	Raw Score	Theta	Scale Score
0	-5.3883	100	25	0.0996	235
1	-4.2427	100	26	0.2473	240
2	-3.6056	100	27	0.4004	246
3	-3.2335	113	28	0.5604	252
4	-2.9626	122	29	0.7282	258
5	-2.7433	130	30	0.9046	264
6	-2.5540	137	31	1.0902	271
7	-2.3836	144	32	1.2849	278
8	-2.2256	149	33	1.4881	286
9	-2.0761	155	34	1.6989	293
10	-1.9323	160	35	1.9163	301
11	-1.7925	165	36	2.1395	309
12	-1.6555	170	37	2.3687	318
13	-1.5203	175	38	2.6045	326
14	-1.3865	180	39	2.8480	335
15	-1.2534	185	40	3.1005	345
16	-1.1208	190	41	3.3634	354
17	-0.9884	195	42	3.6388	364
18	-0.8559	200	43	3.9301	375
19	-0.7232	205	44	4.2460	387
20	-0.5900	209	45	4.6061	400
21	-0.4559	214	46	5.0591	400
22	-0.3206	219	47	5.7665	400
23	-0.1833	224	48	6.9503	400
24	-0.0436	229			

**Raw, Theta, and Scale Scores for Reading Breach**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.3922	100	25	0.1720	237
1	-4.2384	100	26	0.3142	243
2	-3.5946	100	27	0.4604	248
3	-3.2180	113	28	0.6114	253
4	-2.9435	123	29	0.7681	259
5	-2.7206	131	30	0.9314	265
6	-2.5277	138	31	1.1018	271
7	-2.3534	145	32	1.2800	278
8	-2.1913	151	33	1.4664	285
9	-2.0372	156	34	1.6612	292
10	-1.8886	162	35	1.8648	299
11	-1.7438	167	36	2.0776	307
12	-1.6018	172	37	2.3002	315
13	-1.4616	177	38	2.5335	324
14	-1.3228	183	39	2.7783	333
15	-1.1852	188	40	3.0354	342
16	-1.0485	193	41	3.3055	352
17	-0.9125	198	42	3.5892	363
18	-0.7772	203	43	3.8891	374
19	-0.6425	207	44	4.2125	385
20	-0.5082	212	45	4.5787	399
21	-0.3740	217	46	5.0360	400
22	-0.2395	222	47	5.7459	400
23	-0.1040	227	48	6.9302	400
24	0.0329	232			

**Raw, Theta, and Scale Scores for Writing Hybrid 1**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-4.8749	100	31	0.4463	228
1	-3.6882	100	32	0.5476	232
2	-3.0175	100	33	0.6512	236
3	-2.6310	104	34	0.7574	241
4	-2.3584	115	35	0.8665	245
5	-2.1467	124	36	0.9786	249
6	-1.9722	131	37	1.0940	254
7	-1.8225	137	38	1.2128	259
8	-1.6901	142	39	1.3350	264
9	-1.5703	147	40	1.4606	269
10	-1.4600	151	41	1.5895	274
11	-1.3567	155	42	1.7213	279
12	-1.2588	159	43	1.8559	285
13	-1.1652	163	44	1.9931	290
14	-1.0746	167	45	2.1327	296
15	-0.9862	170	46	2.2747	302
16	-0.8996	174	47	2.4195	307
17	-0.8139	177	48	2.5672	313
18	-0.7288	181	49	2.7188	319
19	-0.6439	184	50	2.8750	326
20	-0.5589	188	51	3.0375	332
21	-0.4733	191	52	3.2079	339
22	-0.3870	194	53	3.3891	346
23	-0.2998	198	54	3.5851	354
24	-0.2115	202	55	3.8023	363
25	-0.1221	205	56	4.0519	373
26	-0.0313	209	57	4.3554	385
27	0.0609	212	58	4.7625	400
28	0.1546	216	59	5.4367	400
29	0.2500	220	60	6.6085	400
30	0.3472	224			

**Raw, Theta, and Scale Scores for Writing Hybrid 2**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-4.9290	100	31	0.4274	227
1	-3.6997	100	32	0.5313	231
2	-2.9863	100	33	0.6373	236
3	-2.5738	107	34	0.7458	240
4	-2.2868	118	35	0.8566	245
5	-2.0683	127	36	0.9700	249
6	-1.8922	134	37	1.0858	254
7	-1.7442	140	38	1.2043	258
8	-1.6159	145	39	1.3254	263
9	-1.5018	150	40	1.4490	268
10	-1.3981	154	41	1.5748	273
11	-1.3023	158	42	1.7028	279
12	-1.2124	161	43	1.8326	284
13	-1.1268	165	44	1.9640	289
14	-1.0445	168	45	2.0969	294
15	-0.9643	171	46	2.2312	300
16	-0.8855	174	47	2.3672	305
17	-0.8074	178	48	2.5053	311
18	-0.7294	181	49	2.6464	316
19	-0.6510	184	50	2.7919	322
20	-0.5717	187	51	2.9437	328
21	-0.4910	190	52	3.1050	335
22	-0.4088	194	53	3.2798	342
23	-0.3247	197	54	3.4751	350
24	-0.2384	200	55	3.7019	359
25	-0.1500	204	56	3.9795	370
26	-0.0593	208	57	4.3437	385
27	0.0336	211	58	4.8686	400
28	0.1288	215	59	5.7432	400
29	0.2262	219	60	7.1053	400
30	0.3257	223			

**Raw, Theta, and Scale Scores for Writing Breach**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-4.6557	100	31	0.4345	228
1	-3.4706	100	32	0.5356	232
2	-2.8069	100	33	0.6389	236
3	-2.4304	112	34	0.7446	240
4	-2.1695	123	35	0.8528	244
5	-1.9700	131	36	0.9637	249
6	-1.8080	137	37	1.0772	253
7	-1.6707	143	38	1.1937	258
8	-1.5506	148	39	1.3130	263
9	-1.4430	152	40	1.4351	268
10	-1.3445	156	41	1.5598	273
11	-1.2529	160	42	1.6870	278
12	-1.1664	163	43	1.8163	283
13	-1.0836	166	44	1.9475	288
14	-1.0036	170	45	2.0805	294
15	-0.9254	173	46	2.2152	299
16	-0.8484	176	47	2.3516	305
17	-0.7719	179	48	2.4903	310
18	-0.6953	182	49	2.6321	316
19	-0.6183	185	50	2.7782	322
20	-0.5403	188	51	2.9305	328
21	-0.4611	191	52	3.0921	334
22	-0.3804	195	53	3.2673	341
23	-0.2979	198	54	3.4629	349
24	-0.2136	201	55	3.6898	358
25	-0.1272	205	56	3.9674	370
26	-0.0388	208	57	4.3317	384
27	0.0517	212	58	4.8573	400
28	0.1443	216	59	5.7341	400
29	0.2390	220	60	7.0982	400
30	0.3357	224			

**Raw, Theta, and Scale Scores for Writing Combo1**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-4.9663	100	31	0.4177	227
1	-3.7344	100	32	0.5206	231
2	-3.0154	100	33	0.6256	235
3	-2.5965	106	34	0.7327	240
4	-2.3036	117	35	0.8421	244
5	-2.0802	126	36	0.9537	248
6	-1.9000	134	37	1.0676	253
7	-1.7490	140	38	1.1840	258
8	-1.6183	145	39	1.3028	262
9	-1.5025	150	40	1.4240	267
10	-1.3976	154	41	1.5475	272
11	-1.3010	158	42	1.6731	277
12	-1.2105	161	43	1.8007	282
13	-1.1248	165	44	1.9300	288
14	-1.0423	168	45	2.0609	293
15	-0.9622	171	46	2.1932	298
16	-0.8837	174	47	2.3268	304
17	-0.8061	178	48	2.4620	309
18	-0.7287	181	49	2.5992	315
19	-0.6508	184	50	2.7392	320
20	-0.5723	187	51	2.8834	326
21	-0.4924	190	52	3.0337	332
22	-0.4111	194	53	3.1931	338
23	-0.3278	197	54	3.3659	345
24	-0.2426	200	55	3.5591	353
25	-0.1550	204	56	3.7847	362
26	-0.0652	207	57	4.0650	374
27	0.0270	211	58	4.4512	389
28	0.1214	215	59	5.1098	400
29	0.2181	219	60	6.2772	400
30	0.3168	223			

**Raw, Theta, and Scale Scores for Writing Combo2**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-5.0698	100	31	0.4494	228
1	-3.8532	100	32	0.5545	232
2	-3.1511	100	33	0.6621	237
3	-2.7428	100	34	0.7725	241
4	-2.4550	111	35	0.8859	246
5	-2.2324	120	36	1.0024	250
6	-2.0498	128	37	1.1224	255
7	-1.8940	134	38	1.2458	260
8	-1.7568	139	39	1.3725	265
9	-1.6332	144	40	1.5023	270
10	-1.5197	149	41	1.6350	276
11	-1.4137	153	42	1.7703	281
12	-1.3133	157	43	1.9079	287
13	-1.2172	161	44	2.0475	292
14	-1.1243	165	45	2.1894	298
15	-1.0336	168	46	2.3339	304
16	-0.9445	172	47	2.4815	310
17	-0.8564	176	48	2.6331	316
18	-0.7687	179	49	2.7902	322
19	-0.6811	183	50	2.9544	329
20	-0.5931	186	51	3.1279	336
21	-0.5044	190	52	3.3134	343
22	-0.4151	193	53	3.5146	351
23	-0.3246	197	54	3.7369	360
24	-0.2331	201	55	3.9887	370
25	-0.1402	204	56	4.2847	382
26	-0.0460	208	57	4.6529	397
27	0.0497	212	58	5.1544	400
28	0.1469	216	59	5.9695	400
29	0.2458	220	60	7.2778	400
30	0.3466	224			

**Raw, Theta, and Scale Scores for Writing Combo3**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-4.7029	100	31	0.4251	227
1	-3.5101	100	32	0.5252	231
2	-2.8363	100	33	0.6274	235
3	-2.4516	111	34	0.7319	239
4	-2.1842	122	35	0.8386	244
5	-1.9798	130	36	0.9478	248
6	-1.8141	137	37	1.0595	253
7	-1.6741	143	38	1.1739	257
8	-1.5520	148	39	1.2909	262
9	-1.4429	152	40	1.4106	267
10	-1.3435	156	41	1.5329	272
11	-1.2513	160	42	1.6577	277
12	-1.1643	163	43	1.7847	282
13	-1.0815	167	44	1.9138	287
14	-1.0015	170	45	2.0446	292
15	-0.9235	173	46	2.1772	298
16	-0.8469	176	47	2.3113	303
17	-0.7708	179	48	2.4471	308
18	-0.6948	182	49	2.5850	314
19	-0.6184	185	50	2.7257	320
20	-0.5412	188	51	2.8705	326
21	-0.4628	191	52	3.0213	332
22	-0.3829	195	53	3.1812	338
23	-0.3013	198	54	3.3543	345
24	-0.2178	201	55	3.5478	353
25	-0.1322	205	56	3.7734	362
26	-0.0446	208	57	4.0535	373
27	0.0452	212	58	4.4391	389
28	0.1370	216	59	5.0966	400
29	0.2310	219	60	6.2626	400
30	0.3270	223			

**Raw, Theta, and Scale Scores for Writing Combo4**

<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>	<b>Raw Score</b>	<b>Theta</b>	<b>Scale Score</b>
0	-4.8351	100	31	0.4561	228
1	-3.6542	100	32	0.5583	233
2	-2.9906	100	33	0.6630	237
3	-2.6101	105	34	0.7706	241
4	-2.3424	116	35	0.8813	245
5	-2.1348	124	36	0.9954	250
6	-1.9637	131	37	1.1129	255
7	-1.8167	137	38	1.2341	260
8	-1.6866	142	39	1.3589	265
9	-1.5687	147	40	1.4872	270
10	-1.4597	151	41	1.6188	275
11	-1.3576	155	42	1.7533	281
12	-1.2606	159	43	1.8905	286
13	-1.1674	163	44	2.0302	292
14	-1.0769	167	45	2.1723	297
15	-0.9886	170	46	2.3171	303
16	-0.9016	174	47	2.4652	309
17	-0.8154	177	48	2.6175	315
18	-0.7297	181	49	2.7753	322
19	-0.6440	184	50	2.9402	328
20	-0.5581	188	51	3.1145	335
21	-0.4715	191	52	3.3007	343
22	-0.3843	195	53	3.5027	351
23	-0.2962	198	54	3.7257	360
24	-0.2071	202	55	3.9783	370
25	-0.1168	205	56	4.2750	382
26	-0.0252	209	57	4.6440	397
27	0.0677	213	58	5.1463	400
28	0.1621	217	59	5.9626	400
29	0.2582	220	60	7.2721	400
30	0.3561	224			