

Maryland High School Assessment Program

**2003
Technical Report**

**Submitted to the
Maryland State Department of Education**

February 29, 2004



**CTB/McGraw-Hill
20 Ryan Ranch Road
Monterey, California 93940**

Developed and published by CTB/McGraw-Hill LLC, a subsidiary of The McGraw-Hill Companies, Inc., 20 Ryan Ranch Road, Monterey, California 93940-5703. Copyright © 2004 by Maryland State Department of Education. All rights reserved. No part of this publication may be reproduced or distributed in any form or by any means, or stored in a database or retrieval system, without the prior written permission of Maryland State Department of Education and the publisher.

Table of Contents

Background and Overview	3
January and May 2003 Administrations	4
The 2003 Test Forms	5
Samples for Research Analyses, Calibration, and Equating	12
Initial Data Preparation, Review, and Quality Assurance Procedures	14
Classical Item Analysis and Data Quality Assurance	16
Item-total correlations.	21
Differential Item Functioning.	26
Total raw score distributions and test reliability.	29
Calibration, Equating and Scoring Procedures	37
IRT Model	37
Items Excluded from the Calibrations	38
January 2003 Calibration and Equating Procedures	40
May 2003 Calibration and Equating Procedures	42
Item Fit Assessment	42
Scoring the January and May 2003 Test Forms	43
Scale Score Results and Standard Errors of the Equated Maryland High School Assessments	44
References	59
APPENDIX A	60

Tables

Table 1a. January 2003 Test Forms	5
Table 1b. May 2003 Test Forms	5
Table 2a. Number of Items in Each January Test Form	6
Table 2b. Number of Items in Each May Test Form	7
Table 3a. Percentages of January Students in Selected Categories*	12
Table 3b. Percentages of May Students in Selected Categories*	13
Table 4a. January Case Counts by Content Area and Sample	14
Table 4b. May Case Counts by Content Area and Sample	14
Table 5b. Distribution of Mean Percent Correct for May Selected-Response Items*	17
Table 6a. Distribution of Mean Percent Correct for January Student-Produced Response (SPR) Items	18
Table 6b. Distribution of Mean Percent Correct for May Student-Produced Response (SPR) Items	18
Table 7a. Distribution of Mean Percent of Total Possible Number of Points for January Brief Constructed-Response Items	19
Table 7b. Distribution of Mean Percent of Total Possible Number of Points for May Brief Constructed-Response Items	19
Table 8a. Distribution of Mean Percent of Total Possible Number of Points for January Extended Constructed-Response Items	20
Table 8b. Distribution of Mean Percent of Total Possible Number of Points for May Extended Constructed-Response Items	20
Table 9a. Percent Distributions of Item-Total Correlations for January Selected-Response Items	22

Table 9b. Percent Distributions of Item-Total Correlations for May Selected-Response Items	22
Table 10a. Percent Distributions of Item-Total Correlations for January Student Produced Response Items (SPRs)	23
Table 10b. Percent Distributions of Item-Total Correlations for May Student Produced Response Items (SPRs)	23
Table 11a. Distributions of Item-Total Correlations for January Brief Constructed-Response Items (BCRs)	24
Table 11b. Distributions of Item-Total Correlations for May Brief Constructed-Response Items (BCRs)	24
Table 12a. Distributions of Item-Total Correlations for January Extended Constructed-Response Items (ECRs)	25
Table 12b. Distributions of Item-Total Correlations for May Extended Constructed-Response Items (ECRs)	25
Table 13a. Number of January Items Flagged for DIF Against the Group using Mantel Haenszel (SR/SPR items) and Mantel (CR items) Procedures ^a	27
Table 13b. Number of Items Flagged for DIF Against the Group using Mantel Haenszel (SR/SPR items) and Mantel (CR items) Procedures ^a	28
Table 14a. January Total Raw Score Distributions and Test Reliability	29
Table 14b. May Raw Score Distributions and Test Reliability	30
Table 15a. Summary of January Test Completion Rates.	33
Table 15b. Summary of May Test Completion Rates	34
Table 16a. Items excluded from the January 2003 Calibration.....	39
Table 16b. Items excluded from the May 2003 Calibration	39
Table 17a. January 2003 Scale Scores after Stocking & Lord and Linear Equipercentile Approximation.	44
Table 17b. May 2003 Scale Scores after Stocking & Lord and Linear Equipercentile Approximation.	45
Table 18a. Scale Score Performance by Content Area, January 2002 and January 2003	47
Table 18b. Scale Score Performance by Content Area, May 2002 and May 2003	47
Table 19a. Conditional Standard Errors at Proficiency Cut Points, January 2003	48
Table 19b. Conditional Standard Errors at Proficiency Cut Points, May 2003	48
Table A1. Summary Statistics for January Constructed Response Items.....	61
Table A2. Summary Statistics for May Constructed Response Items	67

Figures

Figure 1a. Frequency Distributions for 2003 Algebra Assessments.	49
Figure 1b. Frequency Distributions for 2003 Biology Assessments.	50
Figure 1c. Frequency Distributions for 2003 English Assessments.	51
Figure 1d. Frequency Distributions for 2003 Geometry Assessments.	52
Figure 1e. Frequency Distributions for 2003 Government Assessments.....	53
Figure 2a. Conditional Standard Errors for 2003 Algebra Assessments	54
Figure 2b. Conditional Standard Errors for 2003 Biology Assessments.....	55
Figure 2c. Conditional Standard Errors for 2003 English Assessments	56
Figure 2d. Conditional Standard Errors for 2003 Geometry Assessments	57
Figure 2e. Conditional Standard Errors for 2003 Government Assessments.....	58

2003 Maryland High School Assessment

Background and Overview

The Maryland High School Assessment (MHSA) was designed to evaluate the academic proficiency of Maryland High School students in English, Algebra, Geometry, Government, and Biology. The items in these tests include selected response (SR), brief constructed response (BCR), extended constructed response (ECR) and, for Algebra and Geometry, student produced response (SPR, also known as grid response items [GR]). Each test is designed to measure Maryland's *Core Learning Goals and Skills for Success*. Detailed information about these goals and skills and about the specific structure and content of each of the high school assessments is available online at www.mdk12.org/mspp/high_school.

Statewide field testing of items for these tests began in May 2000, and has continued each May and January thereafter. Test items were written to meet detailed test blueprints, and field test forms were constructed using the preliminary specifications that had been set for the future operational forms with respect to content, number and types of items. The field test forms were then administered to large numbers of Maryland high school students using test administration procedures similar to those that later would be employed with the operational test. Field test forms were spiraled within classrooms and schools to try to ensure that randomly equivalent groups of students completed each test version.

Beginning with the May 2000 field test, all items have been calibrated with a mixed IRT model using CTB's proprietary Pardux software.¹ After every field test administration, all items within a content area were calibrated together. Anchor items, selected to be representative of the content of the entire test, were distributed throughout each test form. Within each content area, the same set of anchor items appeared in each version and in approximately the same item positions. These anchor items were used to place all items on the base-year scale defined by items in the May 2000 test administration, through the use of a Stocking and Lord equating procedure.² All items that survived the scrutiny of the

¹ A 3-parameter logistic (3PL) model has been used for selected-response items, and a 2-parameter partial credit (2PPC) model has been used for constructed response items.

² With few exceptions, the same set of anchor items was used from year to year.

field test analyses were included in a pool of items from which future operational test forms would be created.

The scale that was established in May 2000 was set to an arbitrary mean of 500 and standard deviation of 50. However, after the January 2002 test administration, the items in each content area were placed on a new scale with a mean of approximately 400 and standard deviation of approximately 40 for each content area.

Following the May 2002 administration, the highest obtainable scale score (HOSS) and lowest obtainable scale score (LOSS) for each content area were set based upon the scale score distributions and conditional standard errors for each of the test forms administered in January or May 2002. The final HOSS and LOSS values are listed below:

Content	LOSS	HOSS
Algebra	240	625
Biology	260	650
English 1	240	625
Geometry	275	575
Government	260	650

In July 2003, CTB conducted a Bookmark standard-setting workshop in Maryland to set proficiency cut scores for these tests. Passing scores were set for Algebra, Biology, English 1 and Government, and three proficiency levels were defined for Geometry:

Content	Scale Score Ranges for Proficiency Levels		
	Basic	Proficient/Passing	Advanced
Geometry	275 to 410	411 to 446	447 to 575
Algebra	240 to 411	412 to 625	--
Biology	260 to 399	400 to 650	--
English 1	240 to 406	407 to 625	--
Government	260 to 393	394 to 650	--

January and May 2003 Administrations

The 2003 Test Forms

The operational test forms for 2003 were constructed using CTB's proprietary ItemSys software (Burket, 1988). This software allows test developers to use IRT item parameters to create parallel forms by matching test characteristic curves (TCCs) and conditional standard errors (CSEMs), while at the same time meeting content and format specifications for each new form. Thirty-five different test forms were administered in January 2003 and 71 were administered in May 2003, as shown in Tables 1a and 1b.

Table 1a. January 2003 Test Forms

	English I	Biology	Geometry	Government	Algebra	Total
Regular Forms	A,B,C,W	A,B,C,W	A,B,C,W	A,B,C,D,W	A,B,W	20 Forms
# to be calibrated & equated	4	4	4	5	3	20 Forms
Makeup Forms*	X,Y	X,Y	X,Y	X,Y	X,Y	10 Forms
Braille Forms	A (1 form)	A (1 form)	A (1 form)	A (1 form)	A (1 form)	5 Forms
Total	6	6	6	7	5	35 Forms

Table 1b. May 2003 Test Forms

	English I	Biology	Geometry	Government	Algebra	Total
Regular Forms	D,E,F G,H,J K,L (8 forms)	E,F,G H,J,K L (7 forms)	D,E,F G,H,JK,L (8 forms)	E,F,G H,J,K L (7 forms)	C,D,E F,G,H J,K,L (9 forms)	39 Forms
Block Field Test Forms	M,N,P (3 forms)	M,N,PQ,R (5 forms)	M (1 form)	M,N,PQ,R,S (6 forms)	M,N (2 forms)	17 Forms
# to be calibrated & equated	11 Forms	12 Forms	9 Forms	13 Forms	11 Forms	56 Forms
Makeup Forms*	X,Y (2 forms)	X,Y (2 forms)	X,Y (2 forms)	X,Y (2 forms)	X,Y (2 forms)	10 Forms
Braille Forms	A (1 form)	A (1 form)	A (1 form)	A (1 form)	A (1 form)	5 Forms
Total	14 Forms	15 Forms	12 Forms	16 Forms	14 Forms	71 Forms

* Because Makeup forms X and Y are identical to Regular Forms A and B, they were scored using the A and B item parameters and were not separately calibrated.

The numbers and types of items in each form are shown in Tables 2a and 2b, below.

Table 2a. Number of Items in Each January Test Form

Content Area	Form Prefix	Item Type**	SR	SPR	BCR	ECR	
Algebra	A & X	FT	6	2	1	1	
		OP	26	6	3	3	
	B & Y	FT	6	2	1	1	
		OP	26	6	3	3	
	W	OP	32	8	4	4	
	Biology	A & X	FT	10	-	2	-
OP			48	-	7	-	
B & Y		FT	10	-	2	-	
		OP	48	-	7	-	
C		FT	10	-	2	-	
		OP	48	-	7	-	
D		FT	10	-	2	-	
		OP	48	-	7	-	
W		OP	58	-	9	-	
English		A & X	FT	15	-	1	-
			OP	50	-	2	1
		B & Y	FT	16	-	1	-
	OP		50	-	2	1	
	C	FT	15	-	1	-	
		OP	50	-	2	1	
	W	OP	70	-	3	1	
	Geometry	A & X	FT	8	2	1	1
OP			26	6	2	3	
B & Y		FT	8	2	1	1	
		OP	26	6	2	3	
C		FT	8	2	1	1	
		OP	26	6	2	3	
W		OP	33	8	3	4	
Government		A & X	FT	6	-	4	-
	OP		50	-	7	1	
	B & Y	FT	6	-	4	-	
		OP	50	-	7	1	
	C	FT	6	-	4	-	
		OP	50	-	7	1	
	D	FT	6	-	4	-	
		OP	50	-	7	1	
	W	OP	74	-	8	1	

** FT=Field Test, OP=Operational, ANC=Anchor

Table 2b. Number of Items in Each May Test Form

Content Area	Form Prefix	Item Type**	SR	SPR	BCR	ECR
Algebra	C	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
	D	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
	E	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
	F	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
	G	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
	H	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
	J	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
	K	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
	L	ANC	11	6	-	-
		OP	15	-	3	3
		FT	6	2	1	1
M	OP	26	6	3	3	
	FT	10	-	1	1	
N	OP	26	6	3	3	
	FT	6	2	1	1	

** FT=Field Test, OP=Operational, ANC=Anchor

Table 2b. Number of Items in Each May Test Form (Continued)

Content Area	Form Prefix	Item Type**	SR	SPR	BCR	ECR
Biology	E	ANC	29	-	-	-
		OP	19	-	7	-
		FT	10	-	2	-
	F	ANC	29	-	-	-
		OP	19	-	7	-
		FT	10	-	2	-
	G	ANC	29	-	-	-
		OP	19	-	7	-
		FT	10	-	2	-
	H	ANC	29	-	-	-
		OP	19	-	7	-
		FT	10	-	2	-
	J	ANC	29	-	-	-
		OP	19	-	7	-
		FT	10	-	2	-
	K	ANC	29	-	-	-
		OP	19	-	7	-
		FT	10	-	2	-
	L	ANC	29	-	-	-
		OP	19	-	7	-
		FT	10	-	2	-
	M	OP	48	-	7	-
		FT	30	-	-	-
	N	OP	48	-	7	-
		FT	10	-	2	-
	P	OP	48	-	7	-
		FT	10	-	2	-
	Q	OP	48	-	7	-
		FT	10	-	2	-
	R	OP	47	-	7	-
FT		11	-	2	-	
English	D	ANC	33	-	-	-
		OP	17	-	2	1
		FT	19	-	1	-
	E	ANC	33	-	-	-
		OP	17	-	2	1
		FT	16	-	1	-
	F	ANC	33	-	-	-
		OP	17	-	2	1
		FT	17	-	1	-

** FT=Field Test, OP=Operational, ANC=Anchor

Table 2b. Number of Items in Each May Test Form (Continued)

Content Area	Form Prefix	Item Type**	SR	SPR	BCR	ECR	
English (Cont'd)	G	ANC	33	-	-	-	
		OP	17	-	2	1	
		FT	17	-	1	-	
	H	ANC	33	-	-	-	
		OP	17	-	2	1	
		FT	16	-	1	-	
	J	ANC	33	-	-	-	
		OP	17	-	2	1	
		FT	17	-	1	-	
	K	ANC	33	-	-	-	
		OP	17	-	2	1	
		FT	17	-	1	-	
	L	ANC	33	-	-	-	
		OP	17	-	2	1	
		FT	15	-	1	-	
	M	OP	50	-	2	1	
		FT	17	-	1	-	
	N	OP	49	-	3	1	
		FT	18	-	-	-	
	P	OP	50	-	2	1	
		FT	16	-	1	-	
	Geometry	D	ANC	11	6	-	-
			OP	15	-	2	3
			FT	8	2	1	1
E		ANC	11	6	-	-	
		OP	15	-	2	3	
		FT	8	2	1	1	
F		ANC	11	6	-	-	
		OP	15	-	2	3	
		FT	8	2	1	1	
G		ANC	11	6	-	-	
		OP	15	-	2	3	
		FT	8	2	1	1	
H		ANC	11	6	-	-	
		OP	15	-	2	3	
		FT	8	2	1	1	
J		ANC	11	6	-	-	
		OP	15	-	2	3	
		FT	8	2	1	1	

** FT=Field Test, OP=Operational, ANC=Anchor

Table 2b. Number of Items in Each May Test Form (Continued)

Content Area	Form Prefix	Item Type**	SR	SPR	BCR	ECR
Geometry (Cont'd)	K	ANC	11	6	-	-
		OP	15	-	2	3
		FT	8	2	1	1
	L	ANC	11	6	-	-
		OP	15	-	2	3
		FT	8	2	1	1
	M	OP	26	6	2	3
		FT	10	1	1	1
	Govern- ment	E	ANC	29	-	-
OP			21	-	7	1
FT			6	-	4	-
F		ANC	29	-	-	-
		OP	21	-	7	1
		FT	6	-	4	-
G		ANC	29	-	-	-
		OP	21	-	7	1
		FT	6	-	4	-
H		ANC	29	-	-	-
		OP	21	-	7	1
		FT	6	-	4	-
J		ANC	29	-	-	-
		OP	21	-	7	1
		FT	6	-	4	-
K		ANC	29	-	-	-
		OP	21	-	7	1
		FT	6	-	4	-
L		ANC	29	-	-	-
		OP	21	-	7	1
		FT	6	-	4	-
M		OP	50	-	7	1
		FT	24	-	1	-
N		OP	50	-	7	1
		FT	5	-	-	1
P		OP	50	-	8	-
		FT	5	-	1	-
Q		OP	50	-	7	1
		FT	5	-	-	1
R		OP	50	-	7	1
	FT	5	-	-	1	
S	OP	50	-	7	1	
	FT	5	-	-	1	

** FT=Field Test, OP=Operational, ANC=Anchor

Forms A through L in each content area were considered “Regular” forms.³ These forms were all linked through a core of common items. Forms N through S were Block Field Test forms, which had no items in common with the Regular forms and virtually no items in common with each other. Form W was a repeated form that had first been administered in May 2002, and shared no common items with the other 2003 test forms; this form was administered in January to help place the new test forms on the existing operational scale.⁴ In addition, Makeup forms X and Y were both administered in January and May. These forms were identical to Forms A and B, respectively.

For each administration, all test forms except the Braille, Large Print, and Makeup forms were packaged with the test forms spiraled in alternating fashion (i.e., with forms interleaved within each package) so that they would be distributed at equal rates within each classroom across the state. This approach was intended to ensure that all forms was administered to randomly equivalent groups of students, with approximately equal sample sizes for all test forms. The use of randomly equivalent samples helps to ensure that the items will be accurately placed on the score scale. However, as will be discussed later in this report, the spiraling design failed to produce the desired results. It appears that test administrators tended to select the first form(s) in each spiral for special education students, ESL students, and other students requiring special testing accommodations. In addition, Maryland experienced a major snowstorm on the date that was scheduled for the January Geometry testing and more than two thirds of all examinees were tested with a makeup form.

³ Within each content area, Form A was also made available in Braille and Large Print.

⁴ Although the “regular” forms included some items from previous field test administrations, it was felt that these field test parameters were likely to be less than fully representative of student performance in an operational setting. The Form W parameters, on the other hand, had been established in the May 2002 operational administration and were therefore considered reasonable indicators of item performance in the operational context.

Samples for Research Analyses, Calibration, and Equating

Cases were considered valid and included in the research data file based on the criteria listed below. Beginning with the January 2003 administration, students with testing accommodations were included in the research analyses. The resulting samples, by gender, ethnicity, and LEA, are summarized in Tables 3a and 3b. To be valid, cases must have:

- Valid LEA number, or blank
- Valid school number, blank, or non-participating school
- An identifiable content area
- A valid version (form) and answer sheet
- No out-of-range item response
- Appropriate number of reads for CR items⁵
- Not coded as invalid by MI or MSDE

Table 3a. Percentages of January Students in Selected Categories*

	Algebra (n=6,950)	Biology (n=10,650)	English (n=9,488)	Geometry (n=9,124)	Government (n=11,179)
Female	48.4	49.2	49.7	51.4	48.9
Male	51.1	50.3	49.7	48.1	50.6
Gender Not Specified	0.5	0.5	0.4	0.5	0.5
African American	32.1	32.9	28.8	32.3	33.6
American Indian	0.2	0.3	0.3	0.2	0.3
Asian	2.6	2.4	2.5	3.1	2.1
Hispanic	5.4	3.3	2.2	3.5	3.3
White	58.5	60.3	65.0	59.9	59.8
Other Ethnicity	1.3	0.9	1.1	1.0	0.9
Baltimore City	11.4	16.8	19.1	17.3	18.0
Other LEA	88.6	83.2	80.9	82.7	82.0

⁵ All Geometry CR items must be scored by at least two raters. If the first two ratings differ by more than one point, then a third rating is required. In other content areas, only a single rating is required, with approximately 10 percent of responses receiving second reads.

Table 3b. Percentages of May Students in Selected Categories*

	Algebra (n=65,115)	Biology (n=52,989)	English (n=56,914)	Geometry (n=50,468)	Government (n=55,426)
Female	49.5	49.7	49.0	50.7	49.1
Male	49.4	49.5	50.1	48.7	50.1
Gender Not Specified	1.1	0.8	0.9	0.6	0.8
African American	33.2	34.3	35.7	31.6	35.4
American Indian	0.3	0.4	0.4	0.4	0.4
Asian	5.4	5.7	5.3	6.2	5.9
Hispanic	5.3	5.3	4.9	5.1	5.1
White	54.4	53.3	52.4	55.8	52.2
Other Ethnicity	1.4	1.1	1.2	0.9	1.1
Baltimore City	7.5	5.5	5.8	6.0	4.7
Other LEA	92.5	94.5	94.2	94.0	95.3

* Includes Regular, Braille, Large Print, and Makeup forms

The samples shown in Tables 3a and 3b included Regular forms, Block Field Test forms, Large Print, Braille, and Makeup forms. In January, students taking Large Print and Braille forms were included in the item analysis and calibration samples for Form A. However, With the exception of the January Geometry test, students taking makeup forms were not included in the item analysis, calibration, or equating. (For reasons discussed later in this report, special administration conditions required that all forms of the January Geometry test be included in the calibrations.)

Because the makeup forms in each content area were identical to the Regular forms A and B, the makeup forms were scored using the item parameters for forms A and B.

Because the Large print and Braille forms administered in May were identical to those administered in January, these forms were not calibrated or equated in May. Therefore, students taking these forms were not included in the May item analysis, calibration, and equating samples. Students who attempted fewer than five items were excluded from the item analysis and calibration/equating samples in January and May.

The resulting January and May case counts are provided in Tables 4a and 4b.

Table 4a. January Case Counts by Content Area and Sample

	Case Counts			
	Total	Regular, Block FT, Braille & Lg Print Examinees	Makeup Form Examinees*	Item Analysis/ Calibration/Equating Sample
Algebra	6,950	5,629	1,321	5,576
Biology	10,650	9,741	909	9,705
English	9,488	8,555	933	8,465
Geometry	9,124	2,982	6,142	8,745
Government	11,179	10,460	719	10,303

* The Makeup forms in each content area were identical to Regular forms A and B and were not included in the January 2003 calibrations.

Table 4b. May Case Counts by Content Area and Sample

	Case Counts			
	Total	Regular & Block FT Examinees	Makeup/Braille/LgPrint Examinees*	Item Analysis/ Calibration/Equating Sample
Algebra	65,115	60,717	4,398	60,069
Biology	52,989	49,783	3,206	49,099
English	56,914	53,371	3,543	52,561
Geometry	50,468	45,484	4,984	44,898
Government	55,426	52,212	3,214	51,495

*The Makeup, Braille, and Large Print forms were calibrated following the January 2003 administration, and were not included in the May 2003 analyses.

Initial Data Preparation, Review, and Quality Assurance Procedures

CTB's quality assurance checks began far in advance of the test administration. Research department personnel with training in the five content areas answered all of the test items and verified the selected-response answer keys that had been entered into the scoring system against their own answers to the test questions and against source documents. Parameter files that were to be used for scoring previously administered items/forms were checked by at least two research associates, and all software was tested using carefully designed test decks to ensure the accuracy of the code.

When data files were received from Measurement Incorporated (MI), the selected response items were scored by CTB's Technology Department and the original and scored data were forwarded to our Statistical Analysis Department. The Statistical Analysis

Department then independently rescored the selected response items and compared the results for quality assurance (QA) purposes. No discrepancies were found.

In addition, the Constructed Response items (scored by MI) were checked to verify that each item received the appropriate number of ratings. For CR items receiving more than one rating, the final raw score for each item was obtained by taking the higher of the first two ratings, provided that the first two ratings differed by no more than one point. (Note that this is equivalent to averaging the first two ratings and rounding up to the nearest integer.) If the first two ratings differed by more than one point, then the third rating (presumed to be an “expert rating”) was used as the final raw score for that item.⁶

Once these checks were completed, frequency distributions were produced by test form for each demographic variable, item response, and item score. Each set of frequency distributions was checked independently by two research associates for reasonableness and accuracy before proceeding with any further analyses.

A comparison of the frequency distributions across test forms revealed the following:

- Although test forms were spiraled within the packages that were delivered to test administrators, disproportionately high numbers of special education students were tested with the first form(s) within each spiral. It appeared that many test administrators were pulling the first form from the stack and using it for students requiring special accommodations. In other words, the spiraling failed to produce randomly equivalent groups of examinees for all test forms.
- Fewer than one third of the January Geometry examinees were tested with a Regular form. Most Geometry examinees could not be tested on the originally scheduled date because of a major snowstorm; when the students were finally tested, most were given a Makeup form

⁶ This is the same scoring rule that is used for the Maryland School Assessment Program.

Because of these results, the item calibration and equating design was modified, as discussed later in this report.

Classical Item Analysis and Data Quality Assurance

Within each of the Regular and Block Field Test forms, items were evaluated for difficulty (mean proportion of students answering the item correctly, or mean percent of maximum possible item score), discrimination (item-total correlation and distractor analysis), and differential item functioning (DIF). These analyses are discussed below.

The classical statistics reported herein were computed excluding omitted responses. Because omit rates were quite high for some of the constructed response items, the statistics presented below probably underestimate the difficulty of those items⁷. On the other hand, since lower ability students are more likely to omit items than higher ability students, the reported item-total correlations are probably underestimates of the true correlations⁸.

The reasons for the high rates of omission are unclear. In the early years of this program, omit rates were much higher than their present values, and it is likely that lack of motivation accounted for at least some of this effect during the field-test years. However, now that test scores are reported to parents and schools, motivational factors are less plausible explanations for the current omit rates. It is not clear whether students simply need more time to complete the test. Although most students do reach the end of the test, that does not rule out the possibility that the tests are speeded. Indeed, completing all of the selected-response items before returning to the more time-consuming constructed response items is a reasonable test-taking strategy. On the other hand, the omit rates may also indicate that students below a certain threshold of ability find it very difficult to generate any kind of productive response to many CR items. Item difficulty.

⁷ Complete item-level data, including omit rates, have been provided to MSDE under separate cover.

⁸ Including omissions would be expected to have the opposite effect, overestimating item difficulty and overestimating item-total correlations and test form reliability.

Classical indicators of item difficulty are the mean percent correct (p+) for the multiple-choice items, and the mean percent of the possible score for the constructed-response items. Overall, it is generally considered desirable to have most items in the 30 percent to 70 percent range. The distributions of item difficulties for selected-response items, presented in Tables 5a and 5b, indicate that the Maryland High School Assessments meet that objective.

Table 5a. Distribution of Mean Percent Correct for Jan. Selected-Response Items*

Mean Percent Correct (p+)	Algebra (% of items)	Biology (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>80	8.5	2.8	14.4	5.0	9.3
71 - 80	8.5	11.2	21.8	2.5	13.0
61 - 70	17.1	15.7	22.7	13.7	18.5
51 - 60	15.9	16.5	16.2	25.5	20.4
41 - 50	23.2	19.7	11.1	26.1	23.0
30 - 40	18.3	25.7	11.1	19.9	13.3
<30	8.5	8.4	2.8	7.5	2.6
Total Number of Items	82	249	216	161	270
Mean p+	52.6	50.8	62.0	49.6	57.0
SD p+	17.5	16.6	17.0	15.1	15.5
Min p+	20	14	15	16	21
Max p+	90	91	89	92	93

* Note: For Algebra, Biology, English and Government, the January statistics exclude examinees who took makeup tests. However, because the majority of Geometry examinees were tested with a makeup form, all Geometry examinees were included in the January analyses.

Table 5b. Distribution of Mean Percent Correct for May Selected-Response Items*

Mean Percent Correct (p+)	Algebra (% of items)	Biology (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>80	7.0	6.3	13.1	6.0	9.9
71 - 80	15.0	14.4	20.8	9.0	16.3
61 - 70	25.5	15.3	25.1	13.7	25.7
51 - 60	20.6	21.0	17.3	24.4	24.8
41 - 50	14.3	22.9	13.3	32.1	17.7
30 - 40	15.4	17.2	7.7	12.8	4.5
<30	2.1	3.0	2.7	2.1	1.1
Total Number of Items	286	576	549	234	649
Mean p+	57.9	55.2	62.6	54.1	61.5
SD p+	15.5	16.0	15.6	14.4	13.6
Min p+	21	15	21	20	24
Max p+	94	96	91	90	95

* Note: The May statistics are based on the May Regular and Block Field Test forms only. Makeup, Braille, and large print forms were repeated from January, and were not included in the May analyses.

The item difficulty statistics for the student-produced gridded response items (SPRs) on the Algebra and Geometry tests are presented in Tables 6a and 6b. In both administrations, the Geometry items were substantially more difficult than the Algebra items. The mean percent correct on the Algebra SPRs was 51.9 percent in January and 61.3 percent in May. The corresponding percentages for the Geometry SPRs were 34.0 percent in January and 39.1 percent in May.

Table 6a. Distribution of Mean Percent Correct for January Student-Produced Response (SPR) Items

Mean Percent Correct (p+)	Algebra (% of items)	Geometry (% of items)
>80	5.0	-
71 - 80	10.0	-
61 - 70	20.0	-
51 - 60	5.0	10.5
41 - 50	40.0	23.7
30 - 40	10.0	23.7
<30	10.0	42.1
Total Number of Items	20	38
Mean p+	51.9	34.0
SD p+	17.3	12.3
Min p+	15	13
Max p+	81	56

Table 6b. Distribution of Mean Percent Correct for May Student-Produced Response (SPR) Items

Mean Percent Correct (p+)	Algebra (% of items)	Geometry (% of items)
>80	12.1	-
71 - 80	27.3	-
61 - 70	10.6	5.6
51 - 60	25.8	24.1
41 - 50	12.1	9.3
30 - 40	7.6	27.8
<30	4.5	33.3
Total Number of Items	66	54
Mean p+	61.3	39.1
SD p+	16.7	13.3
Min p+	22	13
Max p+	83	62

The item difficulty statistics for the brief constructed response items (BCR's) are presented in Tables 7a and 7b. The corresponding statistics for the extended constructed response items (ECR's) are presented in Tables 8a and 8b.

Table 7a. Distribution of Mean Percent of Total Possible Number of Points for January Brief Constructed-Response Items

Mean of Percent Of Possible Points	Algebra (% of items)	Biology ^a (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>80	-	-	-	-	-
71 - 80	-	-	-	-	-
61 - 70	-	-	-	30.8	-
51 - 60	20.0	-	100.0	46.2	2.8
41 - 50	50.0	13.5	-	7.7	66.7
30 - 40	20.0	54.1	-	7.7	30.6
<30	10.0	32.4	-	7.7	-
Points per Item	3	4	4	3	4
Total Number of Items	10	37	9	13	36
Mean p+	44.7	32.5	56.3	54.5	43.3
SD p+	8.8	6.5	2.4	13.2	4.7
Min p+	29	21	54	22	31
Max p+	59	43	60	69	51

Note. ^aThere is no distinction between Biology BCR and ECR items. All Biology CR items are included here in Tables 7a and 7b.

Table 7b. Distribution of Mean Percent of Total Possible Number of Points for May Brief Constructed-Response Items

Mean of Percent Of Possible Points	Algebra (% of items)	Biology ^a (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>80	-	-	-	-	-
71 - 80	3.0	-	-	16.7	-
61 - 70	3.0	-	-	66.7	-
51 - 60	54.5	-	78.3	11.1	14.4
41 - 50	33.3	16.9	21.7	-	70.0
30 - 40	3.0	65.1	-	5.6	13.3
<30	3.0	18.1	-	-	2.2
Points per Item	3	4	4	3	4
Total Number of Items	33	83	23	18	90
Mean p+	49.7	35.6	53.5	64.3	45.4
SD p+	10.0	6.2	4.0	7.5	5.5
Min p+	10	21	45	4.0	2.5
Max p+	75	48	59	7.1	5.3

Note. ^aThere is no distinction between Biology BCR and ECR items. All Biology CR items are included here in Tables 7a and 7b.

Table 8a. Distribution of Mean Percent of Total Possible Number of Points for January Extended Constructed-Response Items

Mean of Percent Of Possible Points	Algebra (% of items)	Biology ^a (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>80	-	-	-	-	-
71 - 80	-	-	-	-	-
61 - 70	10.0	-	-	-	-
51 - 60	10.0	-	100.0	-	-
41 - 50	20.0	-	-	10.5	-
30 - 40	20.0	-	-	36.8	100.0
<30	40.0	-	-	52.6	-
Points per Item	4	-	6	4	4
Total Number of Items	10	-	4	19	5
Mean p+	37.5	-	57.5	29.4	33.6
SD p+	14.8	-	1.0	8.6	1.1
Min p+	19	-	56	12	32
Max p+	63	-	58	45	35

Note. ^aThere is no distinction between Biology BCR and ECR items. All Biology CR items were included in Tables 7a and 7b.

Table 8b. Distribution of Mean Percent of Total Possible Number of Points for May Extended Constructed-Response Items

Mean of Percent Of Possible Points	Algebra (% of items)	Biology ^a (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>80	-	-	-	-	-
71 - 80	-	-	-	-	-
61 - 70	30.3	-	-	-	-
51 - 60	6.1	-	100.0	-	-
41 - 50	30.3	-	-	7.4	38.5
30 - 40	33.3	-	-	63.0	61.5
<30	-	-	-	29.6	-
Points per Item	4	-	6	4	4
Total Number of Items	33	-	11	27	13
Mean p+	32.1	-	59.0	33.6	41.2
SD p+	1.7	-	1.3	6.1	4.4
Min p+	31	-	57	24	36
Max p+	37	-	60	45	50

Note. ^aThere is no distinction between Biology BCR and ECR items. All Biology CR items were included in Tables 7a and 7b.

Although the BCR items were generally of moderate difficulty, the Biology BCR items were considerably more difficult than those in the other content areas, with means of less than

36 percent in both administrations. In contrast, the means for the other four content areas ranged from a low of 43.3 percent (for the January Government items) to a high of 64.3 (for the May Geometry items). However, it should be noted that in the Biology tests no distinction was made between BCR and ECR items, and when the Biology results are compared with the ECR items for the other content areas, the results are somewhat less discrepant.

The ECR items on the English test appear to be of moderate difficulty, with means ranging from 57.5 percent in January and 59.0 percent in May. For the other content areas, however, the ECR items tended to be quite difficult, with means ranging from a low of 29.4 (or the January Geometry items) to a high of only 41.2 percent (for the May Government items). However, as noted earlier, these figures probably underestimate the actual difficulty of the items, since omit rates tended to be quite high⁹.

Item-total correlations.

The item-total correlation (i.e., the correlation between the score on an individual item and the total test score excluding that item¹⁰) indicates the degree to which the content measured by each item matches the content being measured by the test as a whole. The distributions of item-total correlations for selected-response and constructed-response items are presented in Tables 9a/9b (SR items), 10a/10b (SPRs), 11a/11b (BCRs) and 12a/12b (ECRs). It is generally desirable that most items have item-total correlations at or above .15.

The item-total correlations for the Maryland High School Assessments meet this standard. For the selected response items, the percentage with correlations greater than or equal to .15 ranges from a low of 89 percent (January Algebra items) to a high of 99.6 percent (May Algebra items). The results are even stronger for the SPR and constructed response items. All item-total correlations are at or above .22 for the SPR items, .39 for the BCR items, and .50 for the ECR items.

⁹ Omissions were treated as incorrect responses in the item calibrations.

¹⁰ The studied item was removed from the total score so as to not artificially inflate the correlation.

Table 9a. Percent Distributions of Item-Total Correlations for January Selected-Response Items

Correlation (<i>r</i>)	Algebra (% of items)	Biology (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>.54	-	-	1.4	3.7	1.1
.45 - .54	1.2	17.7	18.1	25.5	11.5
.35 - .44	35.4	34.5	37.5	31.1	37.4
.25 - .34	35.4	29.3	25.0	28.0	31.1
.15 - .24	17.1	15.7	13.9	9.3	14.4
<.15	11.0	2.8	4.2	2.5	4.4
Total Number of Items	82	249	216	161	270
Mean <i>r</i>	0.30	0.34	0.35	0.38	0.34
SD <i>r</i>	0.09	0.10	0.11	0.10	0.10
Min <i>r</i>	0.08	0.11	-0.05	0.12	0.06
Max <i>r</i>	0.46	0.51	0.56	0.61	0.58

Table 9b. Percent Distributions of Item-Total Correlations for May Selected-Response Items

Correlation (<i>r</i>)	Algebra (% of items)	Biology (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>.54	3.5	0.7	0.4	6.8	4.0
.45 - .54	28.7	25.2	22.0	32.5	33.9
.35 - .44	36.0	36.1	41.7	37.6	39.9
.25 - .34	24.5	27.4	20.0	18.4	17.3
.15 - .24	7.0	8.9	12.4	3.4	4.3
<.15	0.4	1.7	3.5	1.3	0.6
Total Number of Items	286	576	549	234	649
Mean <i>r</i>	0.39	0.37	0.36	0.41	0.41
SD <i>r</i>	0.10	0.09	0.11	0.10	0.09
Min <i>r</i>	0.13	0.02	0.04	0.03	0.08
Max <i>r</i>	0.63	0.58	0.56	0.66	0.60

Table 10a. Percent Distributions of Item-Total Correlations for January Student Produced Response Items (SPRs)

Correlation (<i>r</i>)	Algebra (% of items)	Geometry (% of items)
>.54	10.0	71.1
.45 - .54	30.0	26.3
.35 - .44	35.0	2.6
.25 - .34	20.0	-
.15 - .24	5.0	-
<.15	-	-
Total Number of Items	20	38
Mean <i>r</i>	0.41	0.60
SD <i>r</i>	0.10	0.08
Min <i>r</i>	0.22	0.37
Max <i>r</i>	0.56	0.71

Table 10b. Percent Distributions of Item-Total Correlations for May Student Produced Response Items (SPRs)

Correlation (<i>r</i>)	Algebra (% of items)	Geometry (% of items)
>.54	25.8	92.6
.45 - .54	27.3	7.4
.35 - .44	34.8	-
.25 - .34	12.1	-
.15 - .24	-	-
<.15	-	-
Total Number of Items	66	54
Mean <i>r</i>	0.47	0.61
SD <i>r</i>	0.09	0.05
Min <i>r</i>	0.30	0.47
Max <i>r</i>	0.64	0.71

Table 11a. Distributions of Item-Total Correlations for January Brief Constructed-Response Items (BCRs)

Correlation (<i>r</i>)	Algebra (% of items)	Biology ^a (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>.54	50.0	89.2	88.9	69.2	83.3
.45 - .54	20.0	10.8	11.1	30.8	16.7
.35 - .44	30.0	-	-	-	-
.25 - .34	-	-	-	-	-
.15 - .24	-	-	-	-	-
<.15	-	-	-	-	-
Total Number of Items	10	37	9	13	36
Mean <i>r</i>	0.51	0.61	0.61	0.61	0.60
SD <i>r</i>	0.07	0.07	0.05	0.08	0.05
Min <i>r</i>	0.39	0.46	0.53	0.51	0.49
Max <i>r</i>	0.57	0.75	0.68	0.73	0.67

Table 11b. Distributions of Item-Total Correlations for May Brief Constructed-Response Items (BCRs)

Correlation (<i>r</i>)	Algebra (% of items)	Biology ^a (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>.54	66.7	86.7	82.6	94.4	87.8
.45 - .54	6.1	13.3	17.4	5.6	12.2
.35 - .44	27.3	-	-	-	-
.25 - .34	-	-	-	-	-
.15 - .24	-	-	-	-	-
<.15	-	-	-	-	-
Total Number of Items	33	83	23	18	90
Mean <i>r</i>	0.55	0.63	0.60	0.63	0.62
SD <i>r</i>	0.09	0.07	0.04	0.07	0.06
Min <i>r</i>	0.39	0.49	0.50	0.53	0.48
Max <i>r</i>	0.65	0.76	0.67	0.72	0.76

Note. ^aThere is no distinction between Biology BCR and ECR items. All Biology CR items are included here in Tables 11a and 11b.

Table 12a. Distributions of Item-Total Correlations for January Extended Constructed-Response Items (ECRs)

Correlation (<i>r</i>)	Algebra (% of items)	Biology ^a (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>.54	70.0	-	75.0	89.5	100.0
.45 - .54	30.0	-	25.0	10.5	-
.35 - .44	-	-	-	-	-
.25 - .34	-	-	-	-	-
.15 - .24	-	-	-	-	-
<.15	-	-	-	-	-
Total Number of Items	10	-	4	19	5
Mean <i>r</i>	0.58	-	0.58	0.61	0.67
SD <i>r</i>	0.05	-	0.05	0.06	0.02
Min <i>r</i>	0.51	-	0.52	0.50	0.64
Max <i>r</i>	0.66	-	0.64	0.78	0.69

Table 12b. Distributions of Item-Total Correlations for May Extended Constructed-Response Items (ECRs)

Correlation (<i>r</i>)	Algebra (% of items)	Biology ^a (% of items)	English (% of items)	Geometry (% of items)	Government (% of items)
>.54	78.8	-	63.6	100.0	100.00
.45 - .54	21.2	-	36.4	-	-
.35 - .44	-	-	-	-	-
.25 - .34	-	-	-	-	-
.15 - .24	-	-	-	-	-
<.15	-	-	-	-	-
Total Number of Items	33	-	11	27	13
Mean <i>r</i>	0.61	-	0.55	0.63	0.66
SD <i>r</i>	0.08	-	0.03	0.04	0.03
Min <i>r</i>	0.53	-	0.52	0.57	0.59
Max <i>r</i>	0.74	-	0.61	0.76	0.70

Note. ^aThere is no distinction between Biology BCR and ECR items. All Biology CR items were included in Tables 11a and 11b.

Differential Item Functioning.

Differential item functioning (DIF) refers to whether different groups perform differently on an item after controlling for total test scores. To evaluate differential item functioning, the Mantel-Haenszel (c^2_{MH}) statistic was used for selected response items and the Mantel statistic was used for constructed response items.

Selected response items were flagged for DIF if the Mantel-Haenszel statistic was significantly greater than zero (at the .05 level) and $|\Delta_{MH}|$ exceeded 1.5. (Zwick, 1993). For the constructed-response items, an effect size (ES) statistic based on Mantel c^2 was used. ES is obtained by dividing the standardized mean difference (SMD) statistic by the standard deviation of the item. A detailed description of these procedures can be found in Zwick, Donoghue, & Grima(1993).

DIF analyses were conducted for all operational items on forms A through L and W, and for the operational and embedded field test items on the block field test forms N through S. Comparisons were made between female and male students (focal group=females; reference group=males), and between African American, Asian, Hispanic and White students (focal groups = African American, Asian, and Hispanic; reference group=White).

The minimum sample sizes and flagging criteria for each procedure are summarized below. These criteria were established during the first year of the Maryland High School program and maintained throughout the course of the contract.

	Mantel-Haenszel SR	Mantel CR
Minimum N per focal group	200	500
Flagging criteria	$\chi^2 p < .05$ & $ \Delta_{MH} > 1.5$	$\chi^2 p < .05$ & $ ES > .25^*$

Items were flagged for DIF only if they were more difficult for the focal groups than would be expected based on the total test scores of students in that subgroup. During the first year of the Maryland High School program, it was decided that items that were less difficult than expected for the focal groups would not be flagged, and this decision was followed throughout the course of the contract.

Tables 13a and 13b show the numbers of items flagged statistically for potential DIF. No Constructed Response items were flagged in either January or May.

Table 13a. Number of January Items Flagged for DIF Against the Group using Mantel Haenszel (SR/SPR items) and Mantel (CR items) Procedures^a

	African American	Asian American	Hispanic American	Female	Male	Total Number Of DIF Flags	Number of Flagged Items ^b
Algebra							
SR	-	-	-	-	-	-	-
SPR	1	-	-	-	-	1	1
BCR	-	-	-	-	-	-	-
ECR	-	-	-	-	-	-	-
Biology							
SR	1	-	-	2	-	3	3
CR	-	-	-	-	-	-	-
English							
SR	1	-	-	-	-	1	1
BCR	-	-	-	-	-	-	-
ECR	-	-	-	-	-	-	-
Geometry							
SR	1	-	-	2	-	3	3
SPR	-	-	-	-	-	-	-
BCR	-	-	-	-	-	-	-
ECR	-	-	-	-	-	-	-
Government							
SR	2	-	-	5	-	7	6
BCR	-	-	-	-	-	-	-
ECR	-	-	-	-	-	-	-

^aThe minimum focal group sample size used was 200 for SR items and 500 for CR items. CR items were flagged if χ^2 $p < .05$ & $SMD \geq 0.25$, where SMD = mean difference divided by total standard deviation.

^bItems flagged for DIF against more than one focal group were included only once in the column total.

Table 13b. Number of Items Flagged for DIF Against the Group using Mantel Haenszel (SR/SPR items) and Mantel (CR items) Procedures^a

	African American	Asian American	Hispanic American	Female	Male	Total Number Of DIF Flags	Number of Flagged Items ^b
Algebra							
SR	-	2	2	4	-	8	8
SPR	-	-	-	-	-	-	-
BCR	-	-	-	-	-	-	-
ECR	-	-	-	-	-	-	-
Biology							
SR	-	3	1	1	-	5	5
CR	-	-	-	-	-	-	-
English							
SR	-	12	8	2	-	22	20
BCR	-	-	-	-	-	-	-
ECR	-	-	-	-	-	-	-
Geometry							
SR	-	3	-	-	-	3	3
SPR	-	-	-	-	-	-	-
BCR	-	-	-	-	-	-	-
ECR	-	-	-	-	-	-	-
Government							
SR	4	20	3	5	-	32	29
BCR	-	-	-	-	-	-	-
ECR	-	-	-	-	-	-	-

^aThe minimum focal group sample size used was 200 for SR items and 500 for CR items. CR items were flagged if χ^2 $p < .05$ & $SMD \geq 0.25$, where SMD = mean difference divided by total standard deviation.

^bItems flagged for DIF against more than one focal group were included only once in the column total.

Of course, differential item functioning is not necessarily a sign of construct-irrelevant item bias against any particular group. The differences may reflect different patterns of ability across different groups, possibly resulting from different curricular experiences for the groups. Therefore, items exhibiting statistical DIF were referred to CTB's content experts, who evaluated the items in consultation with content experts familiar with the culture and educational experiences of the groups in question. Following this review, no items in the January or May assessments were removed from scoring on the basis of statistical DIF.

Total raw score distributions and test reliability.

Tables 14a and 14b show summary statistics for each test form based on total raw scores. These tables also include standard errors of measurement (SEM) and internal consistency reliability estimates (coefficient alpha) for the test forms¹¹.

Table 14a. January Total Raw Score Distributions and Test Reliability

	N Students	Total Possible Score	Mean Score	SD	Raw Scores At These Percentiles					Reliability	SEM
					5 th	25 th	50 th	75 th	95 th		
Algebra											
Form A	2041	50	22.35	9.45	7	15	22	29	38	0.89	3.19
Form B	1807	53	23.98	9.88	7	17	24	31	40	0.89	3.34
Form W	1781	67	26.19	11.75	8	17	26	34	46	0.90	3.72
Biology											
Form A	2251	76	31.18	13.10	14	21	29	40	55	0.92	3.65
Form B	1943	76	32.42	12.66	15	23	31	41	56	0.92	3.65
Form C	1896	76	32.91	13.32	14	23	31	42	57	0.92	3.67
Form D	1825	76	32.85	12.92	15	23	31	42	57	0.92	3.65
Form W	1826	93	35.65	14.31	15	25	34	46	61	0.93	3.78
English											
Form A	2394	64	35.98	12.86	13	27	38	46	54	0.93	3.49
Form B	2111	64	37.52	11.84	15	30	39	46	54	0.91	3.46
Form C	2041	63	38.52	11.82	16	32	41	47	53	0.92	3.32
Form W	2009	85	46.27	14.77	19	37	48	57	67	0.93	3.97
Geometry											
Forms A&X*	5705	50	20.57	11.03	5	12	19	29	40	0.91	3.32
Forms B&Y*	1989	50	18.06	10.90	0	10	15	26	39	0.92	3.14
Form C	716	50	19.14	12.30	0	10	18	29	41	0.93	3.16
Form W	714	64	19.22	13.76	0	9	17	29	44	0.94	3.38
Government											
Form A	2424	82	38.93	14.85	16	28	39	50	63	0.93	3.82
Form B	2063	82	40.33	14.15	17	30	41	51	63	0.93	3.81
Form C	2045	82	40.75	14.22	18	31	41	51	63	0.93	3.82
Form D	1981	82	40.74	13.97	18	31	41	51	63	0.93	3.79
Form W	1947	106	48.50	17.70	19	36	50	61	77	0.94	4.34

As the result of a snowstorm on the date scheduled for the January Geometry testing, the majority of examinees were given a makeup form (X or Y). Because makeup forms X and Y are identical to Regular forms A and B, respectively, the Regular and makeup forms were combined for most analyses.

¹¹ Note that coefficient alpha is considered a lower bound of test reliability (Lord and Novick, 1968).

Table 14b. May Raw Score Distributions and Test Reliability

	N Students	Total Possible Score	Mean Score	SD	Raw Scores At These Percentiles					Reliability	SEM
					5 th	25 th	50 th	75 th	95 th		
Algebra											
Form C	6658	53	25.28	12.10	6	15	25	35	44	0.92	3.40
Form D	5678	53	27.38	11.71	8	18	28	37	45	0.91	3.42
Form E	5614	53	27.33	11.60	8	18	28	37	45	0.91	3.41
Form F	5529	53	27.76	11.59	8	19	28	37	45	0.91	3.38
Form G	5480	53	26.25	11.54	7	17	27	36	44	0.91	3.42
Form H	5464	53	27.20	11.63	8	18	28	37	45	0.91	3.41
Form J	5394	53	26.51	11.61	8	17	27	36	44	0.91	3.41
Form K	5314	53	27.91	11.54	8	19	29	37	45	0.91	3.38
Form L	5260	53	27.54	11.80	8	18	28	37	45	0.92	3.40
Form M	5188	53	25.77	10.66	9	18	26	34	43	0.91	3.26
Form N	5138	53	24.44	11.31	7	16	24	33	43	0.91	3.35
Biology											
Form E	5073	76	32.87	14.73	11	22	32	43	59	0.94	3.62
Form F	4198	76	34.93	13.95	14	25	34	45	59	0.93	3.65
Form G	4175	76	34.94	14.27	14	24	33	46	59	0.93	3.64
Form H	4147	76	34.99	13.87	15	25	34	45	59	0.93	3.64
Form J	4163	76	35.23	14.03	15	25	34	45	60	0.93	3.66
Form K	4114	76	34.35	14.42	13	23	34	45	59	0.94	3.64
Form L	4060	76	33.83	13.76	14	24	32	43	58	0.93	3.68
Form M	4052	72	30.70	11.46	13	23	30	39	50	0.92	3.36
Form N	4026	76	32.16	13.25	14	22	31	42	56	0.93	3.63
Form P	3967	76	33.52	14.07	13	23	33	43	58	0.93	3.72
Form Q	3917	77	31.52	12.79	13	22	30	40	54	0.92	3.67
Form R	3891	75	31.86	13.34	12	22	31	41	55	0.93	3.57
English											
Form D	5927	64	35.87	13.40	13	26	38	47	55	0.93	3.52
Form E	4877	64	38.03	12.61	15	30	40	48	55	0.93	3.41
Form F	4870	64	37.91	11.94	16	30	40	47	54	0.92	3.47
Form G	4855	64	37.47	12.28	16	29	39	47	54	0.92	3.46
Form H	4840	64	37.89	12.35	15	30	40	47	55	0.92	3.43
Form J	4784	64	37.21	12.24	16	29	39	47	54	0.92	3.48
Form K	4745	64	39.39	12.16	17	32	42	49	55	0.92	3.42
Form L	4678	64	38.02	12.58	15	30	40	48	55	0.92	3.46
Form M	4663	64	37.61	11.60	17	31	39	46	54	0.91	3.49
Form N	4576	67	37.79	12.55	16	30	39	47	56	0.92	3.52
Form P	4556	64	35.99	11.78	15	28	37	45	53	0.91	3.56

Table 14b. May Raw Score Distributions and Test Reliability (Continued)

	N Students	Total Possible Score	Mean Score	SD	Raw Scores At These Percentiles					Reliability	SEM
					5 th	25 th	50 th	75 th	95 th		
Geometry											
Form D	5880	50	21.38	12.08	5	11	19	31	43	0.93	3.25
Form E	5125	50	21.94	12.22	6	11	20	32	43	0.93	3.28
Form F	5059	50	22.90	11.72	6	13	21	32	43	0.92	3.28
Form G	5056	50	21.38	11.68	6	12	19	31	42	0.92	3.35
Form H	4968	50	21.95	11.77	6	12	20	31	43	0.92	3.30
Form J	4940	50	21.93	11.45	6	12	20	31	42	0.92	3.31
Form K	4871	50	21.76	11.83	6	12	20	31	42	0.92	3.32
Form L	4844	50	23.08	11.58	7	14	22	32	43	0.92	3.30
Form M	4741	50	20.22	10.51	6	12	19	28	39	0.91	3.14
Government											
Form E	4991	82	39.60	17.05	14	26	40	53	66	0.95	3.78
Form F	4043	82	43.65	15.70	17	32	44	56	68	0.94	3.75
Form G	4096	82	43.37	15.89	17	32	44	56	67	0.94	3.73
Form H	4001	82	44.99	15.52	18	34	47	57	67	0.94	3.68
Form J	4030	82	44.40	16.07	17	33	46	57	68	0.95	3.70
Form K	4004	82	45.14	16.01	18	33	47	58	68	0.95	3.68
Form L	4014	81	43.92	15.33	17	33	45	56	67	0.94	3.68
Form M	3962	78	36.93	14.79	15	26	36	48	61	0.94	3.80
Form N	3906	82	40.10	16.06	15	28	40	53	65	0.95	3.82
Form P	3853	82	39.03	15.82	14	27	39	51	64	0.94	3.80
Form Q	3826	82	42.10	14.91	17	32	43	53	65	0.94	3.72
Form R	3789	82	41.03	15.89	15	29	41	54	65	0.95	3.73
Form S	3697	82	40.20	15.51	15	29	41	52	64	0.94	3.79

Reliability was satisfactory for all forms in both January and May. All forms were moderate to high in difficulty. Reliability was satisfactory for all forms in both January and May. Reliability coefficients ranged from .89 to .94 in January and from .91 to .95 in May, exceeding the .85 threshold that has been judged adequate for high stakes standardized tests (see, e.g., Phillips, 2000).

Table 15a. Summary of January Test Completion Rates.

	Percent Who Answered All Items	Percent Who Answered All SR and SPR Items	Percent Who Answered These Numbers of BCR Items											Percent Who Answered These Numbers of ECR Items						
			0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5	
Algebra																				
Form A	40	62	6	7	19	68	-	-	-	-	-	-	-	-	7	12	25	56	-	-
Form B	46	67	5	7	17	72	-	-	-	-	-	-	-	-	7	12	21	60	-	-
Form W	23	44	7	6	10	21	57	-	-	-	-	-	-	-	8	10	14	24	44	-
Biology																				
Form A	48	86	3	2	2	4	5	13	19	51	-	-	-	-	-	-	-	-	-	-
Form B	47	85	2	1	3	3	6	13	21	51	-	-	-	-	-	-	-	-	-	-
Form C	49	88	3	1	3	3	6	12	20	52	-	-	-	-	-	-	-	-	-	-
Form D	48	85	2	1	2	3	7	14	20	52	-	-	-	-	-	-	-	-	-	-
Form W	45	84	2	2	3	4	4	5	7	10	16	48	-	-	-	-	-	-	-	-
English																				
Form A	75	86	5	10	85	-	-	-	-	-	-	-	-	-	10	90	-	-	-	-
Form B	78	86	4	8	88	-	-	-	-	-	-	-	-	-	9	91	-	-	-	-
Form C	78	86	4	9	88	-	-	-	-	-	-	-	-	-	8	92	-	-	-	-
Form W	63	72	4	3	9	84	-	-	-	-	-	-	-	-	8	92	-	-	-	-
Geometry																				
Forms A&X*	44	58	10	18	73	-	-	-	-	-	-	-	-	-	10	10	19	60	-	-
Forms B&Y*	33	51	17	23	60	-	-	-	-	-	-	-	-	-	18	13	20	48	-	-
Form C	40	59	17	19	65	-	-	-	-	-	-	-	-	-	18	11	15	56	-	-
Form W	31	52	16	7	19	58	-	-	-	-	-	-	-	-	18	7	10	18	47	-
Government																				
Form A	64	91	4	2	2	3	3	6	13	67	-	-	-	-	16	84	-	-	-	-
Form B	64	91	3	2	1	2	4	6	14	68	-	-	-	-	15	85	-	-	-	-
Form C	63	91	3	1	2	2	3	7	15	67	-	-	-	-	15	85	-	-	-	-
Form D	65	91	3	1	2	3	4	6	13	70	-	-	-	-	15	85	-	-	-	-
Form W	55	86	3	1	3	4	4	5	7	13	60	-	-	-	23	77	-	-	-	-

*As the result of a snowstorm on the date scheduled for Geometry testing, the majority of examinees were given a makeup form (X or Y). Because makeup forms X and Y are identical to Regular forms A and B, respectively, the Regular and makeup forms were combined for most analyses.

Table 15b. Summary of May Test Completion Rates (continued).

	Percent Who Answered All Items	Percent Who Answered All SR and SPR Items	Percent Who Answered These Numbers of BCR Items											Percent Who Answered These Numbers of ECR Items					
			0	1	2	3	4	5	6	7	8	9	10	0	1	2	3	4	5
English																			
Form D	74	87	6	10	84	-	-	-	-	-	-	-	-	12	88	-	-	-	-
Form E	79	89	4	8	88	-	-	-	-	-	-	-	-	9	91	-	-	-	-
Form F	80	90	4	8	88	-	-	-	-	-	-	-	-	9	91	-	-	-	-
Form G	79	88	4	8	88	-	-	-	-	-	-	-	-	9	91	-	-	-	-
Form H	79	88	4	8	88	-	-	-	-	-	-	-	-	8	92	-	-	-	-
Form J	80	89	4	8	88	-	-	-	-	-	-	-	-	8	92	-	-	-	-
Form K	80	90	4	8	88	-	-	-	-	-	-	-	-	9	91	-	-	-	-
Form L	77	89	5	8	87	-	-	-	-	-	-	-	-	10	90	-	-	-	-
Form M	78	88	4	6	90	-	-	-	-	-	-	-	-	10	90	-	-	-	-
Form N	73	87	4	5	11	-	-	-	-	-	-	-	-	8	92	-	-	-	-
Form P	75	87	5	8	87	-	-	-	-	-	-	-	-	12	88	-	-	-	-
Geometry																			
Form D	52	65	10	18	72	-	-	-	-	-	-	-	-	11	10	18	61	-	-
Form E	52	64	9	18	73	-	-	-	-	-	-	-	-	10	10	18	62	-	-
Form F	51	65	9	18	73	-	-	-	-	-	-	-	-	11	10	19	60	-	-
Form G	52	63	9	18	73	-	-	-	-	-	-	-	-	10	10	17	62	-	-
Form H	53	65	9	17	74	-	-	-	-	-	-	-	-	10	9	18	63	-	-
Form J	54	67	9	17	75	-	-	-	-	-	-	-	-	9	9	18	63	-	-
Form K	53	65	9	18	73	-	-	-	-	-	-	-	-	10	10	17	63	-	-
Form L	53	66	10	17	73	-	-	-	-	-	-	-	-	11	9	18	62	-	-
Form M	51	62	6	20	74	-	-	-	-	-	-	-	-	12	10	18	60	-	-

Table 15b. Summary of May Test Completion Rates (continued).

	Percent Who Answered All Items	Percent Who Answered All SR and SPR Items	Percent Who Answered These Numbers of BCR Items												Percent Who Answered These Numbers of ECR Items					
			0	1	2	3	4	5	6	7	8	9	10	11	0	1	2	3	4	5
Government																				
Form E	61	92	5	2	2	3	4	6	12	65	-	-	-	-	19	81	-	-	-	-
Form F	65	93	4	2	2	3	3	6	12	69	-	-	-	-	16	84	-	-	-	-
Form G	65	92	4	2	2	2	3	6	11	69	-	-	-	-	16	84	-	-	-	-
Form H	66	93	4	2	2	3	3	6	10	70	-	-	-	-	17	83	-	-	-	-
Form J	65	92	4	2	2	3	3	6	11	69	-	-	-	-	17	83	-	-	-	-
Form K	65	93	3	2	2	2	4	5	11	70	-	-	-	-	17	83	-	-	-	-
Form L	66	93	4	1	2	3	3	6	12	70	-	-	-	-	16	84	-	-	-	-
Form M	60	91	5	3	3	5	7	13	65	-	-	-	-	20	80	-	-	-	-	
Form N	59	92	5	1	2	3	5	8	13	62	-	-	-	-	21	79	-	-	-	-
Form P	55	93	4	2	3	4	5	8	13	61	-	-	-	-	26	74	-	-	-	-
Form Q	64	92	4	1	2	2	4	6	13	67	-	-	-	-	17	83	-	-	-	-
Form R	57	92	4	2	2	3	4	8	14	63	-	-	-	-	28	72	-	-	-	-
Form S	63	92	4	1	2	2	4	5	12	70	-	-	-	-	25	75	-	-	-	-

Calibration, Equating and Scoring Procedures

IRT Model

Student item responses were calibrated using CTB's PARDUX software (Burket, 1991) for the concurrent calibration of selected response (SR) and constructed response (CR) items. In these calibrations, omissions were treated as incorrect responses. A three-parameter logistic model (3PL) was used to scale the SR items, and a two-parameter partial credit (2PPC) model was employed to scale the CR items. A brief explanation of the models is provided below.

The 3PL model (Lord & Novick, 1968; Lord, 1980) defines each SR item in terms of three item parameters: (a) an item discrimination parameter, (b) an item location or difficulty parameter, and (c) a guessing parameter. In this model, the probability that a student with scale score \mathbf{q} responds correctly to item j is

$$p_j(\mathbf{q}) = c_j + \frac{(1 - c_j)}{1 + \exp[-1.7a_j(\mathbf{q} - b_j)]},$$

where a_j is the item discrimination, b_j is the item difficulty, and c_j is the probability of a correct response by a person completely lacking in ability.

The 2PPC model defines each CR item in terms of item discrimination as well as location parameter for each score point. The 2PPC model is a special case of Bock's (1972) nominal model, which states that the probability of an examinee with ability \mathbf{q} having a score at the k -th level of the j -th item is

$$P_{jk}(\mathbf{q}) = P(x_j = k - 1 | \mathbf{q}) = \frac{\exp Z_{jk}}{\sum_{i=1}^{m_j} \exp Z_{ji}}, k = 1, \dots, m_j,$$

where m_j is the number of score levels, and

$$Z_{jk} = A_{jk} \theta + C_{jk},$$

$$A_{jk} = a_j (k-1), \quad k = 1, 2, \dots, m_j, \text{ and}$$

$$C_{jk} = -\sum_{i=0}^{k-1} \gamma_{ji}, \quad \text{where } g_{j0} = 0,$$

where A_{jk} is the discrimination parameter of the k -th category of item j , C_{jk} is the intercept of the nonlinear response function associated with the k -th category of item j , α_j and γ_{ji} are the parameters to be estimated from the data. For each item there are $m_j - 1$ independent γ_{ji} parameters and one α_j parameter; a total of m_j independent item parameters are estimated.

Items Excluded from the Calibrations

In both January and May it was necessary to remove some items from the calibration and equating. In January, one Algebra item was removed (Item 20 on Form A) because it had been publicly released prior to the test administration. In the other content areas, items from the repeated test form (W) were removed if they had not been included in the May 2002 calibrations. Two additional items from Geometry Form W were removed because of unacceptable classical item statistics (negative item-total correlations). Both of these items had acceptable statistics in 2002.

In the May calibrations, 21 items were excluded because of poor classical statistics. In order to meet the content blueprint, these items were replaced with appropriate "embedded field test" items that were administered on those forms.

Three additional items were excluded from the May calibrations because of scoring errors. Two Biology items were not scored at all, and one Geometry item did not receive the second ratings required for all operational Geometry items.

The items excluded from the January and May calibrations are listed in Tables 16a and 16b.

Table 16a. Items excluded from the January 2003 Calibration

Content	Form	Item No.	Reason
Algebra	A	20	Turned off due to Public Release per MSDE
Biology	W	67	Turned off in the May 2002 administration
English	W	25	Turned off in the May 2002 administration
English	W	28	Turned off in the May 2002 administration
English	W	32	Turned off in the May 2002 administration
English	W	36	Turned off in the May 2002 administration
Geometry	W	19	Classical Item Statistics
Geometry	W	29	Classical Item Statistics
Government	W	7	Turned off in the May 2002 administration
Government	W	17	Turned off in the May 2002 administration
Government	W	38	Turned off in the May 2002 administration
Government	W	68	Turned off in the May 2002 administration

Table 16b. Items excluded from the May 2003 Calibration

Content	Form	Item No.	Reason	Replaced with Item
Algebra	N	29	Poor Classical Statistics	N2
Algebra	N	25	Poor Classical Statistics	N35
English	N	3	Poor Classical Statistics	N56
English	N	8	Poor Classical Statistics	N59
Biology	M	12	Poor Classical Statistics	M8
Biology	M	38	Poor Classical Statistics	M46
Biology	M	29	Not scored by MI	--
Biology	P	51	Not scored by MI	--
Biology	N	34	Poor Classical Statistics	N56
Biology	Q	15	Poor Classical Statistics	Q39
Biology	Q	18	Poor Classical Statistics	Q40
Biology	Q	31	Poor Classical Statistics	Q46
Biology	Q	34	Poor Classical Statistics	Q43
Biology	Q	49	Poor Classical Statistics	Q52
Biology	R	26	Poor Classical Statistics	R38
Biology	R	43	Poor Classical Statistics	R36
Government	L	51	Poor Classical Statistics	L63
Government	M	8	Poor Classical Statistics	M82
Government	M	71	Only 10% Second Reads by MI	--
Government	P	11	Poor Classical Statistics	P55
Government	R	42	Poor Classical Statistics	R61
Government	S	36	Poor Classical Statistics	S56
Government	S	26	Poor Classical Statistics	S45
Government	S	55	Poor Classical Statistics	S43

January 2003 Calibration and Equating Procedures

Calibrating and Equating the Algebra, Biology, English, and Government Tests

The original calibration and equating design assumed that all test forms within a content area would be administered to randomly equivalent groups, and test forms were spiraled in order to achieve this goal. In the end, however, it was not possible to achieve this goal, for the following reasons:

1. Large print and Braille forms were available for Form A only, resulting in disproportionate numbers of accommodated students receiving these forms.
2. Special Education students tended to be overrepresented on the first couple of forms within each content area. It appears that administrators tended to use the first one or two forms in each package for a disproportionate number of students who required special accommodations.

Because of the requirement that these students be included in the calibration and equating, it was not possible to sample down in order to achieve comparable groups across test forms. Instead, the following strategy was adopted:

1. Scale scores were computed for the students who took the spiraled form W, using the May 2002 parameters. The January 2003 Form W score distribution was compared with the January and May 2002 scale score distributions. Results were consistent with expectations (i.e., means were comparable to January 2002, but lower than the Form W means in May 2002; standard deviations were similar to the standard deviations in prior administrations.)
2. Within each content area, all items were calibrated together as originally intended, using the 2002 Form W parameters as equating anchors
3. Form W was used to equate the form immediately preceding it in each spiral, using a linear approximation to equipercentile equating. (A comparison of sample

demographics by test form indicated that the students taking this form were very similar to those taking Form W.)

4. The transformation constants from Step 3 were applied to all items, and the transformed parameters were used to score all new forms. Form W was scored using the 2002 item parameters.

Calibrating and Equating the Geometry Test

Unfortunately, Maryland experienced a severe snowstorm on the date scheduled for the geometry test administration, and fewer than 3,000 students completed the test on that date. As a result, the spiraled sample contained only small numbers of students (approximately 700 usable cases per form) from a small number of LEA's. The vast majority of students in the spiraled sample came from Anne Arundel, Frederick, Howard, Washington, and Baltimore City.

More than 6,000 students were tested at a later date with a makeup form. Approximately 4,900 received makeup form X, and approximately 1,200 received makeup form Y.

In light of this situation, the calibration and equating plan for Geometry was modified as follows:

1. Scale scores were computed for the students who took the spiraled form W, using the May 2002 parameters. The January 2003 Form W score distribution was compared with the January and May 2002 Geometry scale score distributions. Results were consistent with expectations (i.e., mean was comparable to January 2002, but lower than the Form W mean in May 2002; standard deviation was similar to the standard deviation in prior administrations.)
2. The 22 core items (i.e., the items common to all forms except Form W) were calibrated together, using the responses of all students in the primary and makeup groups.
3. The parameters for these core items were then fixed, and the remaining operational items were calibrated together.

4. The new tests/items were placed on the current Maryland HS Geometry scale through an equipercentile approximation to the Form W score distribution. Because the demographic characteristics of examinees taking Form W were found to be comparable to the characteristics of students taking the last of the new forms, item parameters for the last new form in the spiral were equated to Form W. For this transformation, the original (May 2002) item parameters were used to score Form W. CTB's WinFlux software was then used to find the linear transformation that best approximated equipercentile score equating. The transformation constants from this procedure were then applied to the item parameters for all of the new test forms.

May 2003 Calibration and Equating Procedures

In May 2003, all new tests with form designations of L or below shared the same core of common items that were included in all of the January 2003 test forms. Because of this strong core, it was possible to equate these new forms and place them on the January 2003 scale by calibrating all forms together and using the January 2003 item parameters for the common items as anchors in a Stocking and Lord equating procedure.

However, because the Block Field Test forms shared no (or almost no) items in common with the Regular forms, this procedure was not sufficient to place these forms on the reporting scale.

To place the Block Field Test forms on scale, each of these forms was equated (using a linear approximation to equipercentile equating) to the last Regular form within each content area.

Item Fit Assessment

For both administrations, a statistical procedure was used to identify items that did not fit the IRT model. Item model fit information was obtained for each item using a Z -statistic. The Z -statistic is a transformation of the chi-square (Q_1) statistic that takes into account differing numbers of score levels as well as sample size:

$$Z_j = \frac{(Q_{1j} - DF_j)}{\sqrt{2DF_j}}$$

where Q_{1j} is the item chi-square statistic,
 j is an item, and
 DF is the degrees of freedom for a given item j .

The Z -statistic is an index of the degree to which obtained proportions of students with each item score are close to the proportions that would be predicted by the estimated student ability and item parameters. These values, along with the associated chi-squares (Q_1) are computed for ten intervals corresponding to deciles of the ability distribution (Yen, 1984). Because the value of Z increases as the sample size increases, with other things being equal, the critical values for Z were established using the following equation (Yen, 1991a):

$$Z_{crit,j} = \frac{4N_j}{1500}$$

where $Z_{crit,j}$ is critical value of Z for item j and
 N_j is the number of students who responded to item j .

Although this criterion did result in the identification of some misfitting items, it was not possible to remove these items without compromising the desired content representation within each form. Detailed information about these items was provided to MSDE in a separate file after completion of the calibrations.

Scoring the January and May 2003 Test Forms

All tests were pattern-scored using the final equated IRT item parameters. Because pattern scoring was used, students obtaining the same raw score on a particular test form usually do not receive the same scale score. Nevertheless, tables of estimated raw-to-scale score values and standard errors were produced for each test form, and were used as initial values in the IRT scoring algorithm. These tables were delivered to MSDE under separate cover.

Scale Score Results and Standard Errors of the Equated Maryland High School Assessments

Tables 17a and 17b show the January and May scale score means and standard deviations for each test form after the Stocking and Lord equating and after the subsequent linear approximation to equipercentile equating.

Table 17a. January 2003 Scale Scores after Stocking & Lord and Linear Equipercentile Approximation.

Content Area	Test Form		After Stocking & Lord		After Linear Equipercentile	
	Form ID	N of Cases	Mean	S.D.	Mean**	S.D.
Algebra	A	2027	399.00	42.71	398.52	43.60
	B	1787	402.84	40.20	402.44	41.06
	W*	1762	402.30	41.42	--	--
English	A	2370	380.75	45.80	390.84	38.07
	B	2090	387.60	41.74	396.42	34.61
	C	2019	386.52	41.62	395.54	34.61
	W*	1986	395.46	34.43	--	--
Biology	A	2239	392.10	41.17	394.83	41.28
	B	1938	395.23	40.76	397.94	40.82
	C	1890	396.23	40.81	398.97	40.85
	D	1819	397.23	39.23	399.97	39.29
	W*	1819	398.79	40.78	--	--
Geometry	A+X	5610**	405.10	39.54	397.74	39.97
	B+Y	1880**	397.59	37.59	389.52	38.61
	C	640	406.03	39.15	398.43	40.46
	W*	629	394.08	41.72	--	--
Government	A	2380	393.75	41.03	395.08	40.48
	B	2038	397.12	39.95	398.39	39.38
	C	2012	397.49	38.45	398.79	37.85
	D	1956	398.05	38.98	399.31	38.34
	W*	1917	399.15	38.34	--	--

* Repeated test forms were scored with their May 2002 item parameters. Note that because an equipercentile approximation was used, the means and standard deviations for Form W do not necessarily match the equated means and standard deviations for the last operational forms in the spiral. Geometry is the most discrepant case because of differently skewed distributions.

** For Geometry, Form A+X includes 740 spiraled forms and 4,870 non-spiraled makeup forms; Form B+Y includes 661 spiraled forms and 1,219 non-spiraled makeup forms.

Table 17b. May 2003 Scale Scores after Stocking & Lord and Linear Equipercentile Approximation.

Content Area	Test Form		After Stocking & Lord		After Linear Equipercentile*	
	Form ID	N of Cases	Mean	S.D.	Mean	S.D.
Algebra	C	6585	405.8	51.4	--	--
	D	5639	413.9	46.0	--	--
	E	5559	414.5	45.9	--	--
	F	5468	414.6	45.0	--	--
	G	5420	412.6	45.9	--	--
	H	5405	413.4	45.3	--	--
	J	5340	414.0	45.5	--	--
	K	5256	414.2	45.6	--	--
	L	5194	413.9	45.0	--	--
	M	5134	410.2	46.0	413.9	44.7
	N	5069	412.2	46.5	414.0	45.0
English	D	5831	390.3	39.8	--	--
	E	4797	397.7	34.5	--	--
	F	4806	398.0	34.8	--	--
	G	4772	397.1	34.3	--	--
	H	4775	397.5	35.5	--	--
	J	4720	397.9	35.5	--	--
	K	4673	399.4	34.4	--	--
	L	4600	398.8	36.2	--	--
	M	4596	388.1	38.7	398.7	36.7
	N	4508	393.0	36.9	398.7	36.7
	P	4483	395.3	37.7	398.9	35.9
Biology	E	4965	399.8	47.0	--	--
	F	4147	405.8	43.6	--	--
	G	4126	406.6	42.9	--	--
	H	4099	406.5	42.1	--	--
	J	4123	406.4	43.6	--	--
	K	4051	405.0	43.9	--	--
	L	4017	405.6	42.3	--	--
	M	4004	401.4	45.0	405.5	42.7
	N	3970	402.8	45.3	405.2	43.3
	P	3910	401.5	44.9	405.4	43.1
	Q	3866	402.2	42.9	405.6	42.3
R	3821	402.6	44.4	405.5	42.9	

* Note that because an equipercentile approximation was used, the means and standard deviations for Form L do not necessarily match the equated means and standard deviations for the block field test forms.

Table 17b. May 2003 Scale Scores after Stocking & Lord and Linear Equipercentile Approximation (Continued).

Content Area	Test Form		After Stocking & Lord		After Linear Equipercentile	
	Form ID	N of Cases	Mean	S.D.	Mean	S.D.
Geometry	D	5790	399.3	43.5	--	--
	E	5069	402.6	42.6	--	--
	F	4993	404.3	40.0	--	--
	G	4997	404.1	41.2	--	--
	H	4912	403.6	42.3	--	--
	J	4874	404.6	39.7	--	--
	K	4813	403.8	41.4	--	--
	L	4782	404.5	39.1	--	--
	M	4668	400.3	40.0	404.4	40.0
Government	E	4897	398.7	49.5	--	--
	F	4005	407.9	45.5	--	--
	G	4037	408.4	46.0	--	--
	H	3957	407.8	44.6	--	--
	J	3971	408.8	46.3	--	--
	K	3954	409.0	45.5	--	--
	L	3977	409.1	45.5	--	--
	M	3900	401.6	43.9	409.2	45.6
	N	3849	403.3	43.5	409.1	45.5
	P	3805	396.3	45.7	408.6	45.9
	Q	3773	400.1	43.9	409.9	45.5
	R	3726	400.7	45.2	410.2	46.0
	S	3644	401.1	44.0	409.8	45.6

Scale score performance across test forms is shown in Table 18a and 18b. Each of these tables also includes the results from the corresponding administration in 2002. Whereas the preceding tables in this report reflected performance of the calibration and equating sample (which excluded those students who attempted fewer than 5 test items), the statistics in Tables 18a and 18b are based on the total population.

Table 18a. Scale Score Performance by Content Area, January 2002 and January 2003

Content	January 2002			January 2003		
	N	Mean	S.D.	N	Mean	S.D.
Algebra	6,509	396.3	42.9	6,850	395.5	47.6
Biology	10,508	396.8	42.0	10,650	395.6	42.9
English	9,339	398.3	41.0	9,488	389.5	42.2
Geometry	9,502	385.6	41.9	9,124	391.5	45.9
Government	11,391	398.0	41.4	11,179	394.9	43.9

Table 18b. Scale Score Performance by Content Area, May 2002 and May 2003

Content	May 2002			May 2003		
	N	Mean	S.D.	N	Mean	S.D.
Algebra	63,877	406.0	52.0	64,700	409.6	49.2
Biology	50,567	399.8	45.0	52,581	401.8	46.3
English	52,172	395.4	47.0	56,426	394.6	39.5
Geometry	45,239	401.0	46.1	50,089	400.1	43.4
Government	52,435	397.7	46.5	54,917	405.2	48.0

Scale score distributions for the January and May 2003 test administrations are shown in Figures 1a through 1e. Conditional standard errors throughout the score range are shown in Figures 2a through 2e. Conditional standard errors at the cut points are listed in Tables 19a and 19b.

Table 19a. Conditional Standard Errors at Proficiency Cut Points, January 2003

Content	Conditional Standard Errors at Cut Points	
	Proficient/Passing	Advanced
Geometry	6.69	6.83
Algebra	8.87	
Biology	8.72	
English 1	8.23	
Government	8.07	

Table 19b. Conditional Standard Errors at Proficiency Cut Points, May 2003

Content	Conditional Standard Errors at Cut Points	
	Proficient/Passing	Advanced
Geometry	7.05	6.80
Algebra	8.97	
Biology	9.04	
English 1	8.40	
Government	8.22	

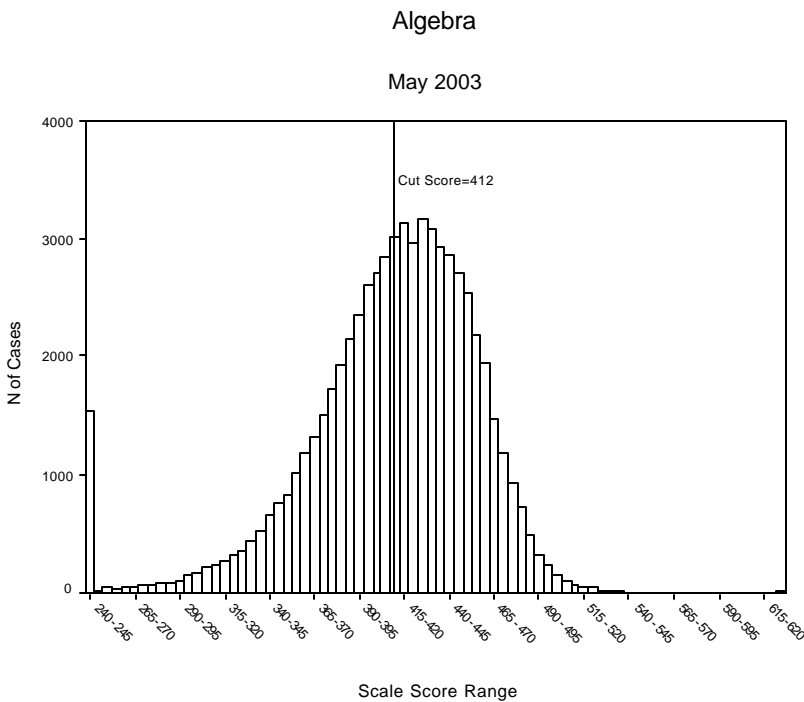
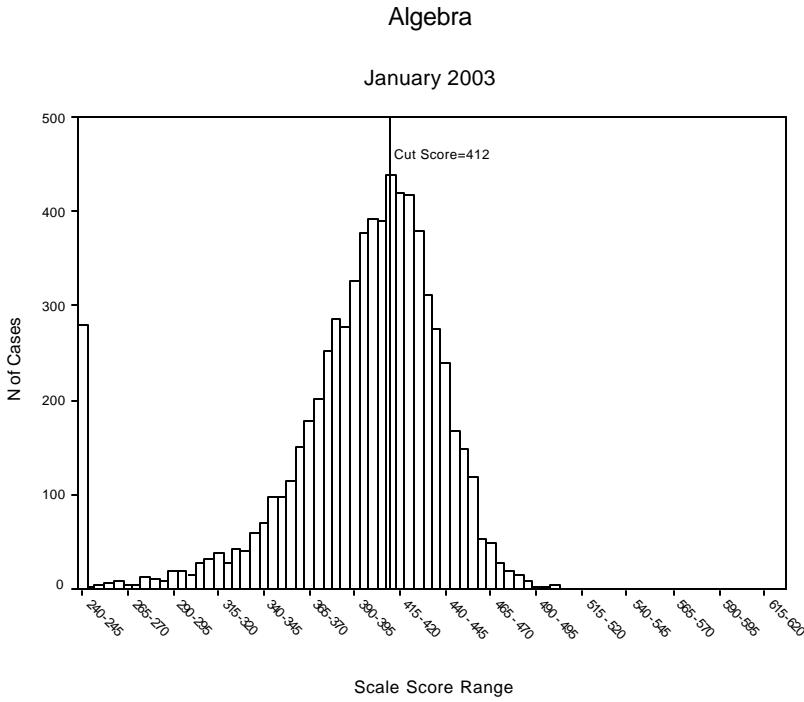


Figure 1a. Frequency Distributions for 2003 Algebra Assessments.

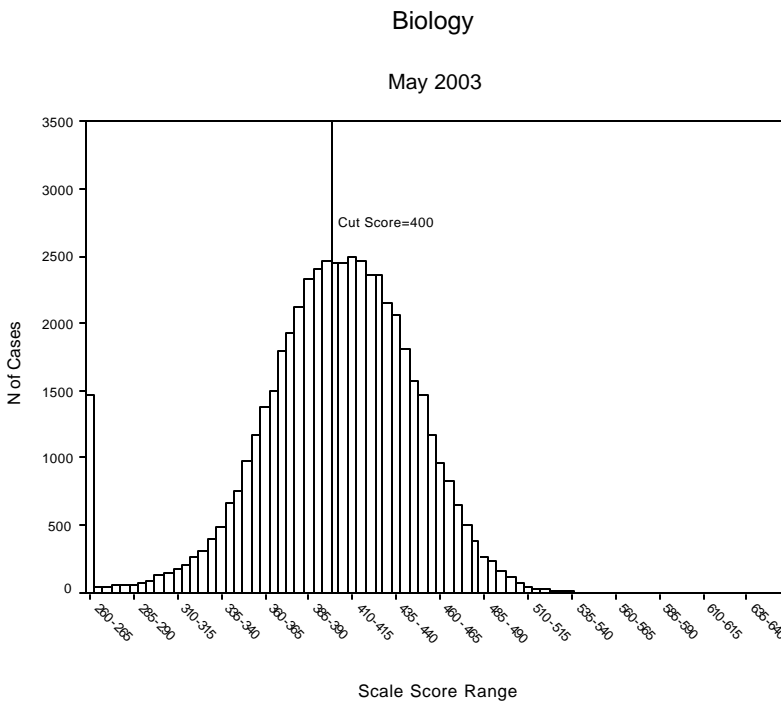
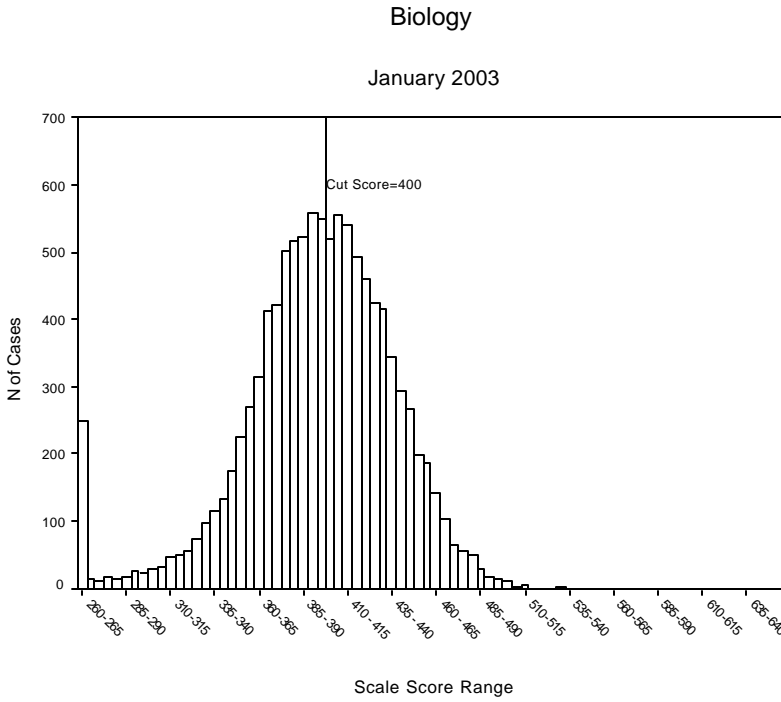


Figure 1b. Frequency Distributions for 2003 Biology Assessments.

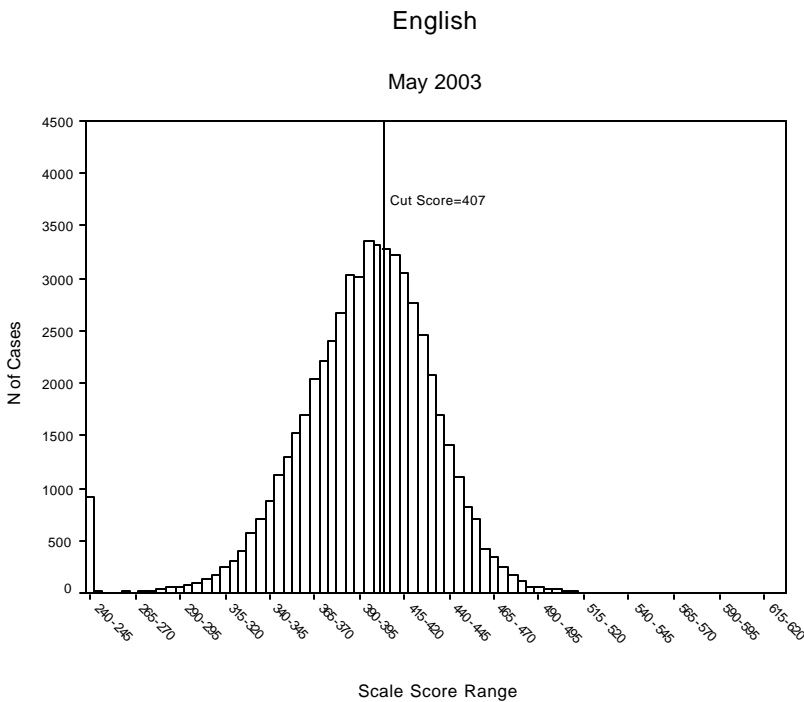
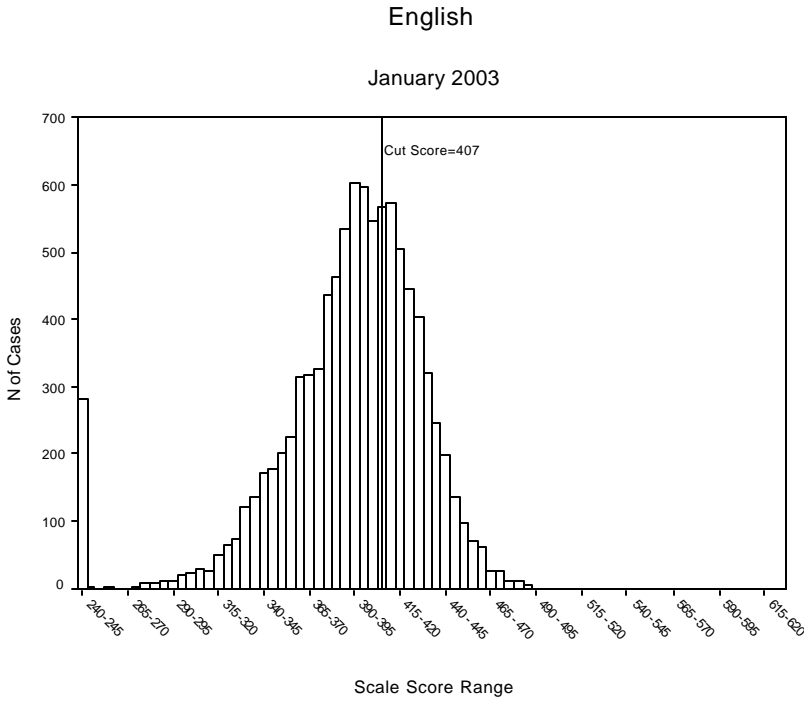


Figure 1c. Frequency Distributions for 2003 English Assessments.

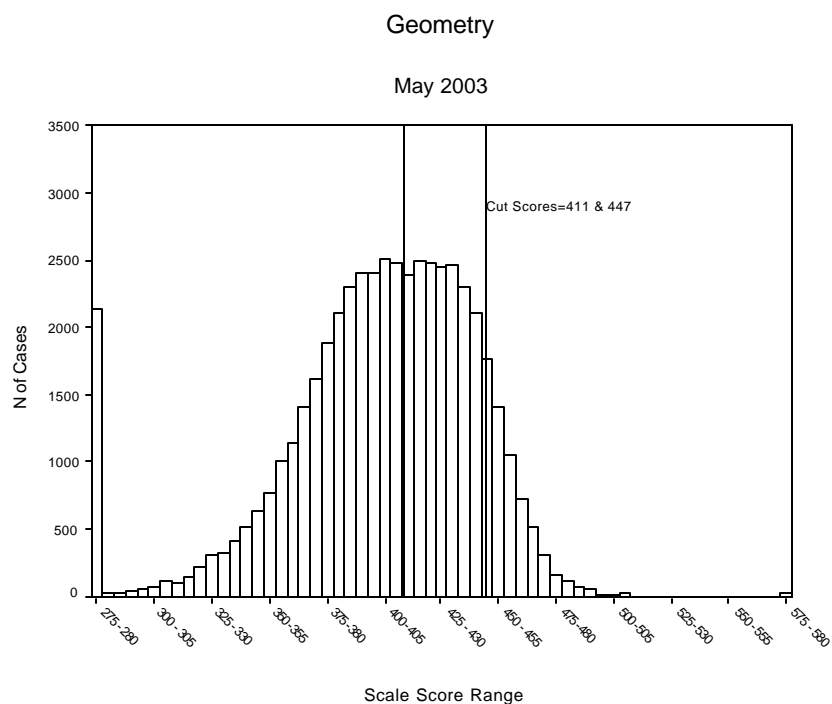
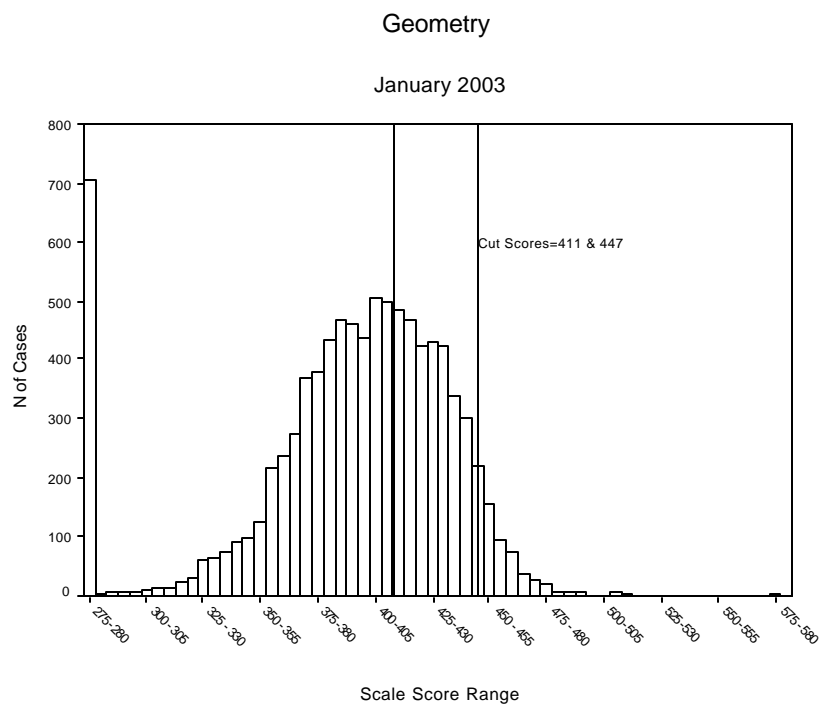


Figure 1d. Frequency Distributions for 2003 Geometry Assessments.

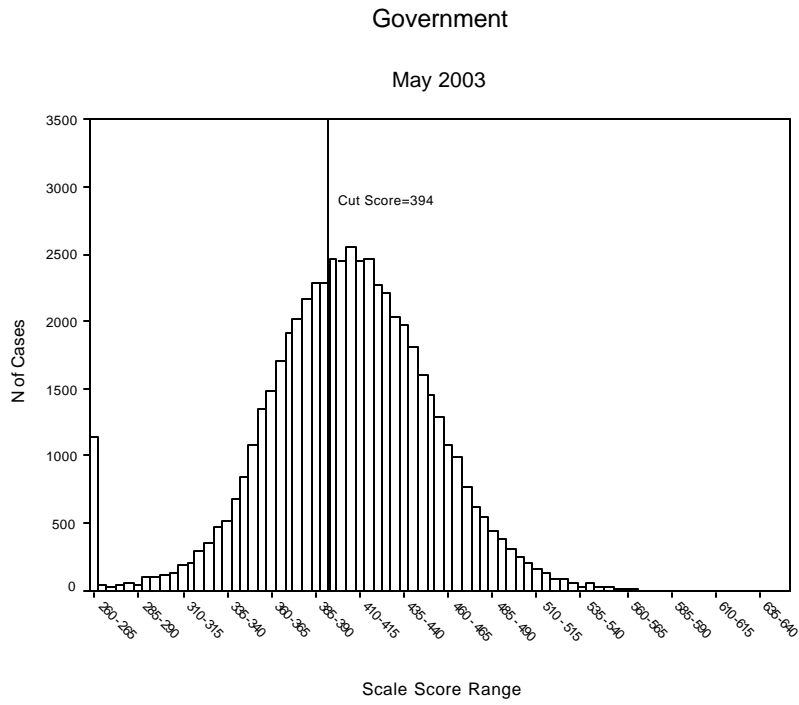
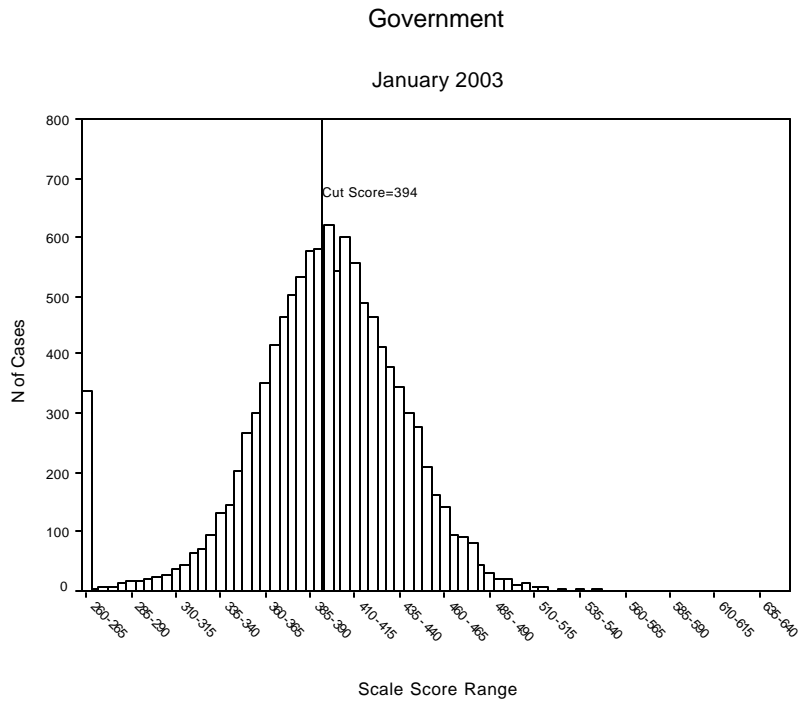


Figure 1e. Frequency Distributions for 2003 Government Assessments.

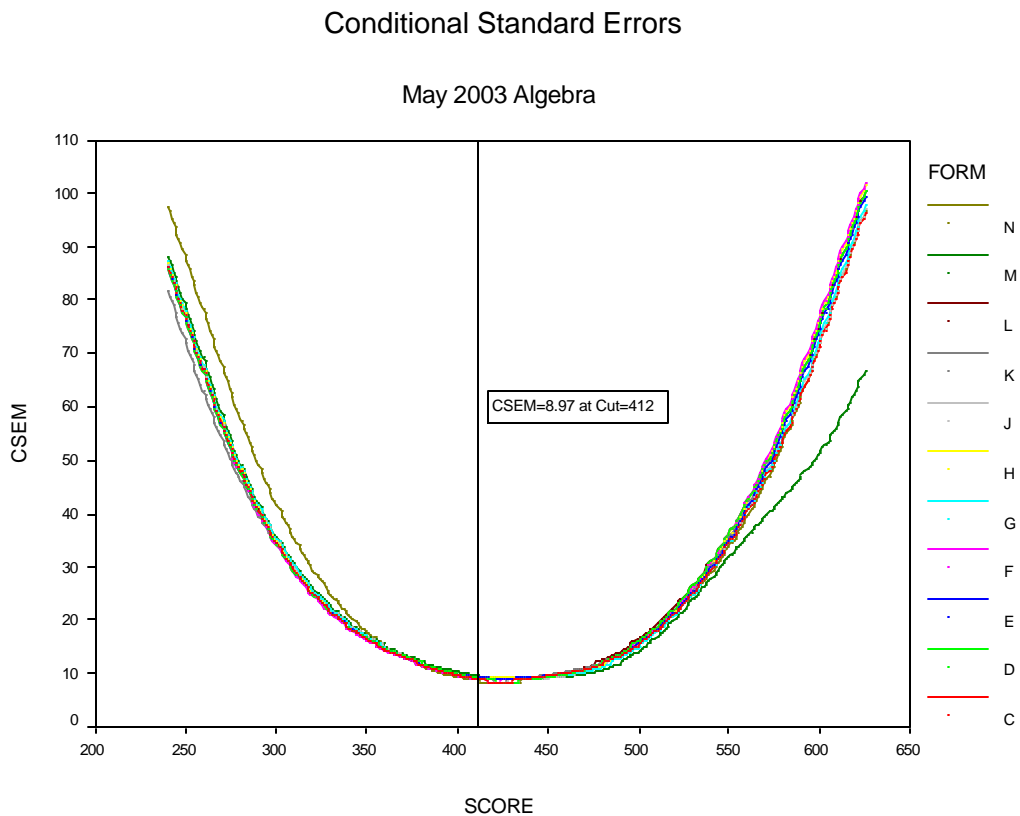
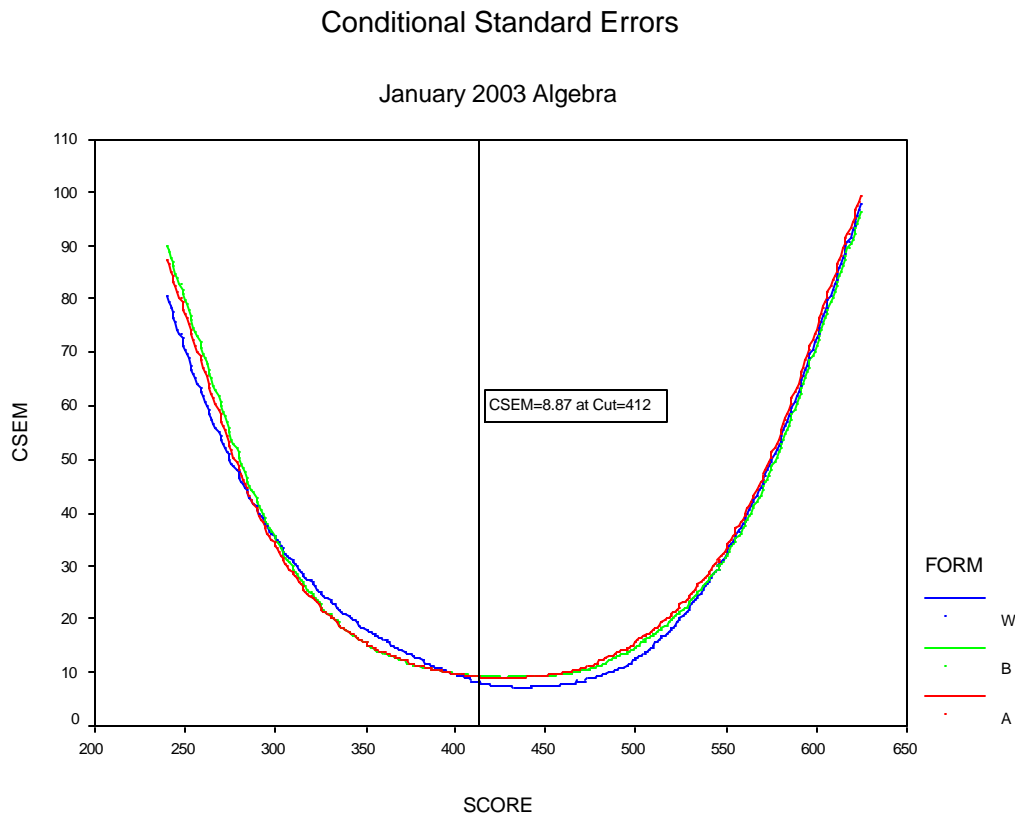
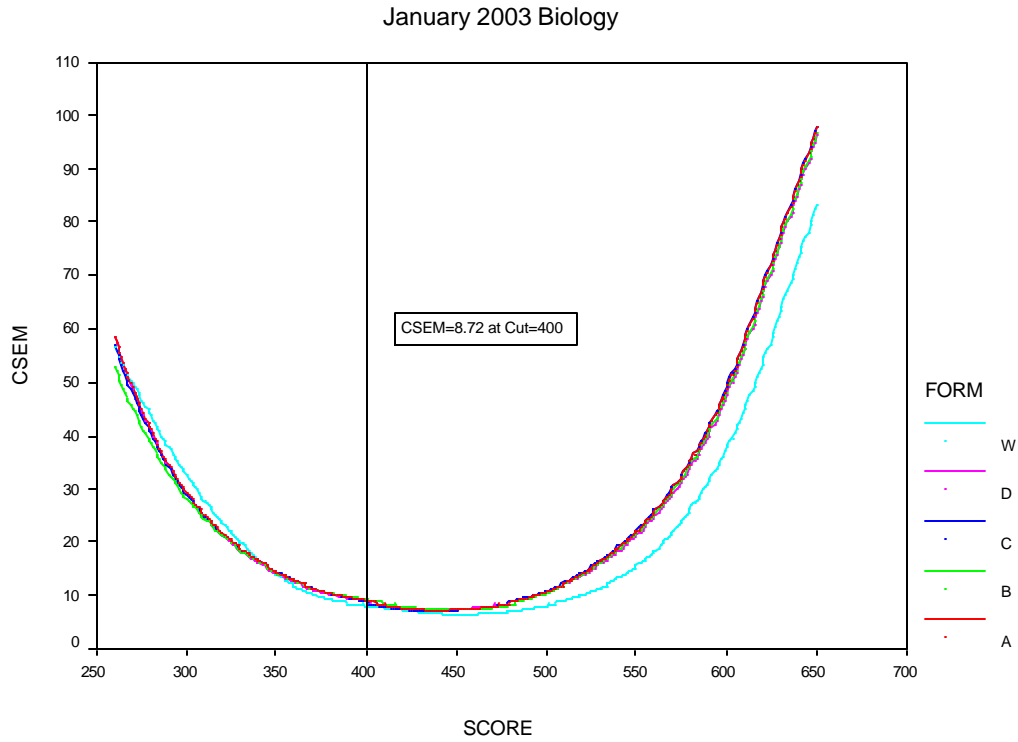


Figure 2a. Conditional Standard Errors for 2003 Algebra Assessments

Conditional Standard Errors



Conditional Standard Errors

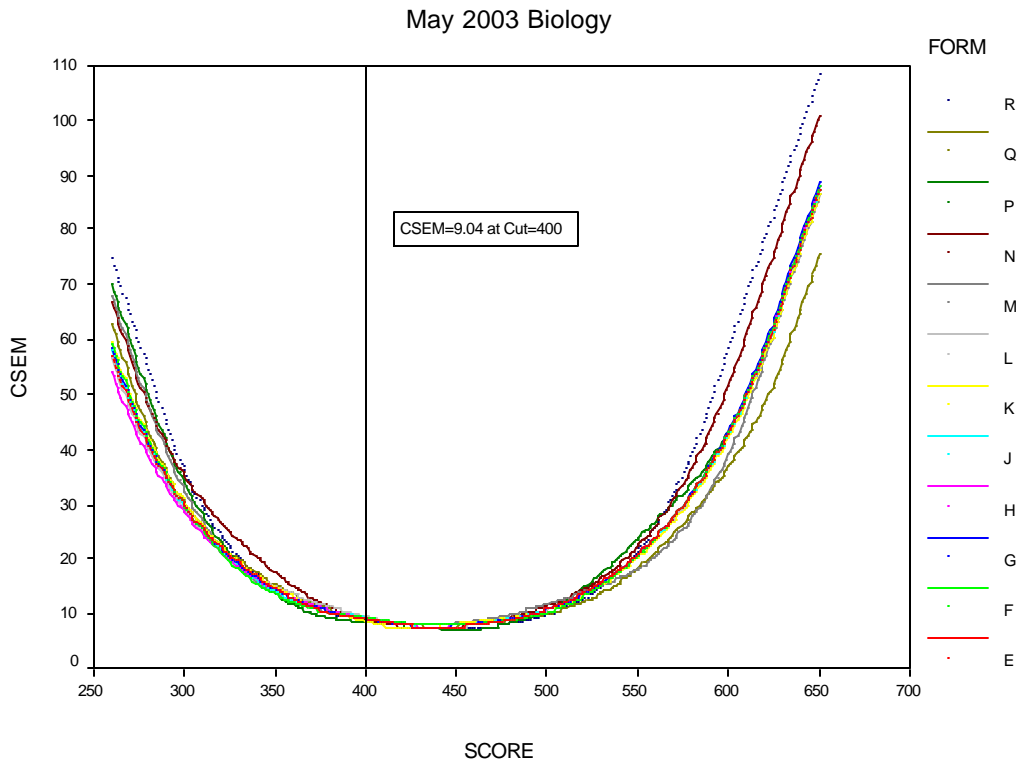
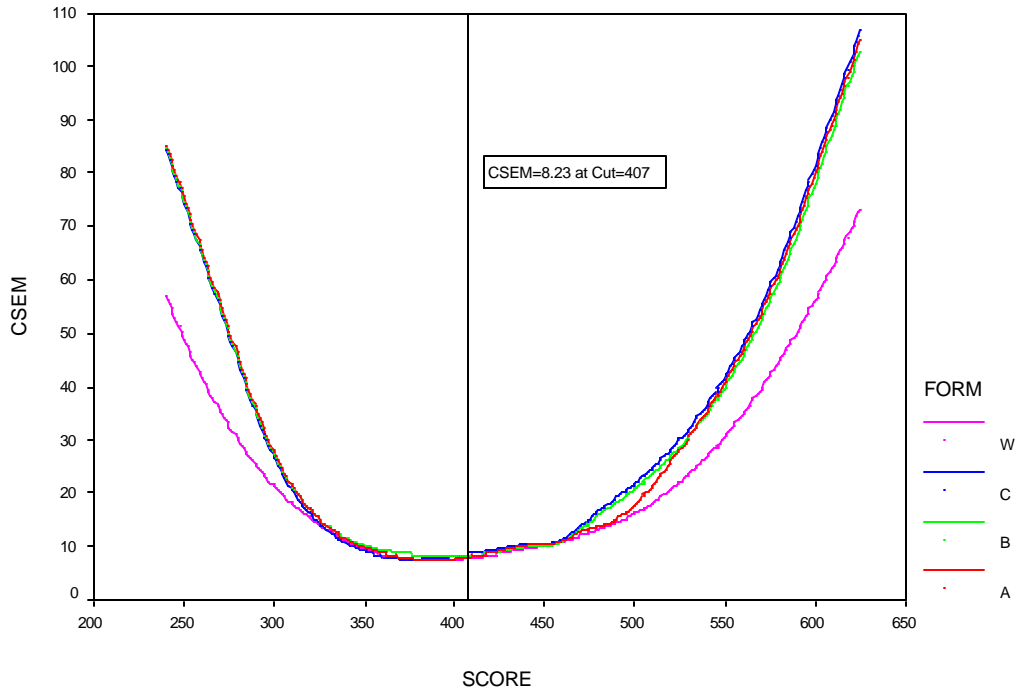


Figure 2b. Conditional Standard Errors for 2003 Biology Assessments

Conditional Standard Errors

January 2003 English 1



Conditional Standard Errors

May 2003 English 1

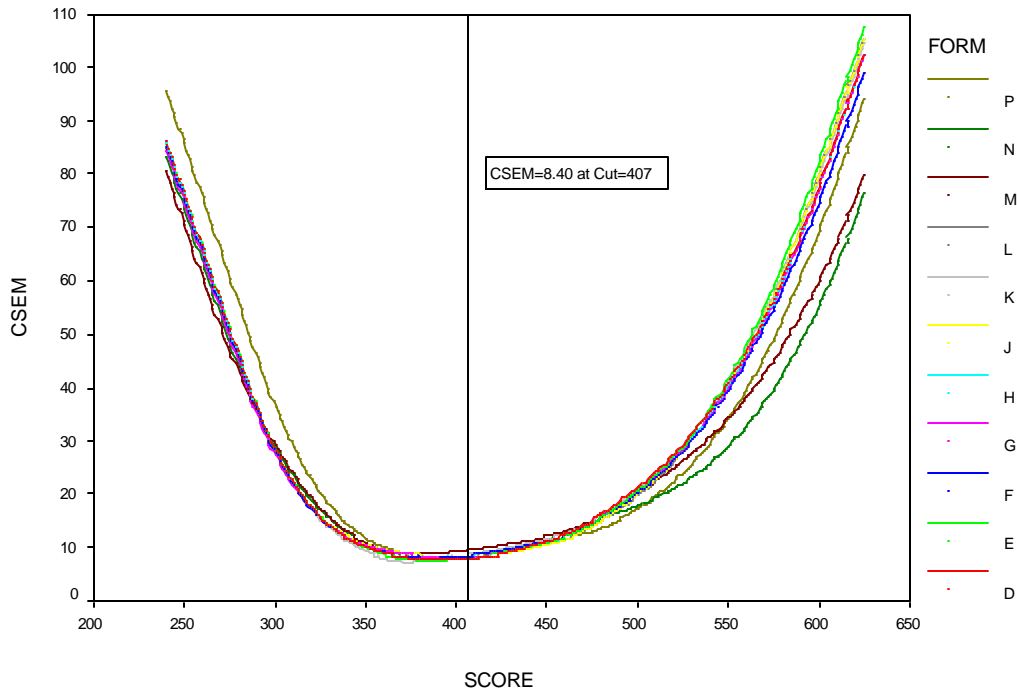
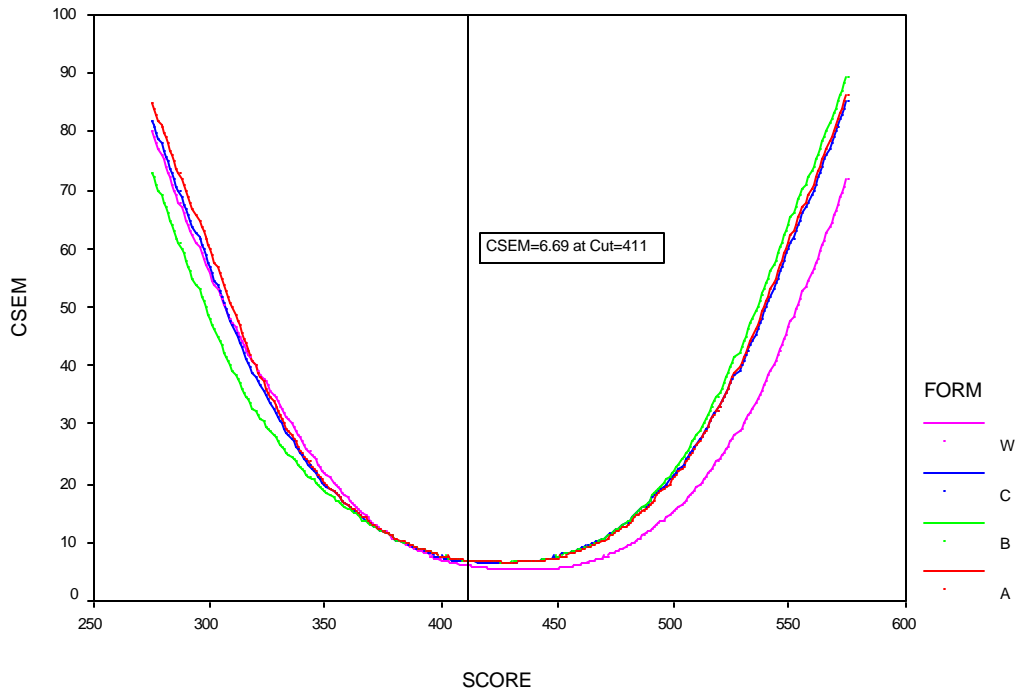


Figure 2c. Conditional Standard Errors for 2003 English Assessments

Conditional Standard Errors

January 2003 Geometry



Conditional Standard Errors

May 2003 Geometry

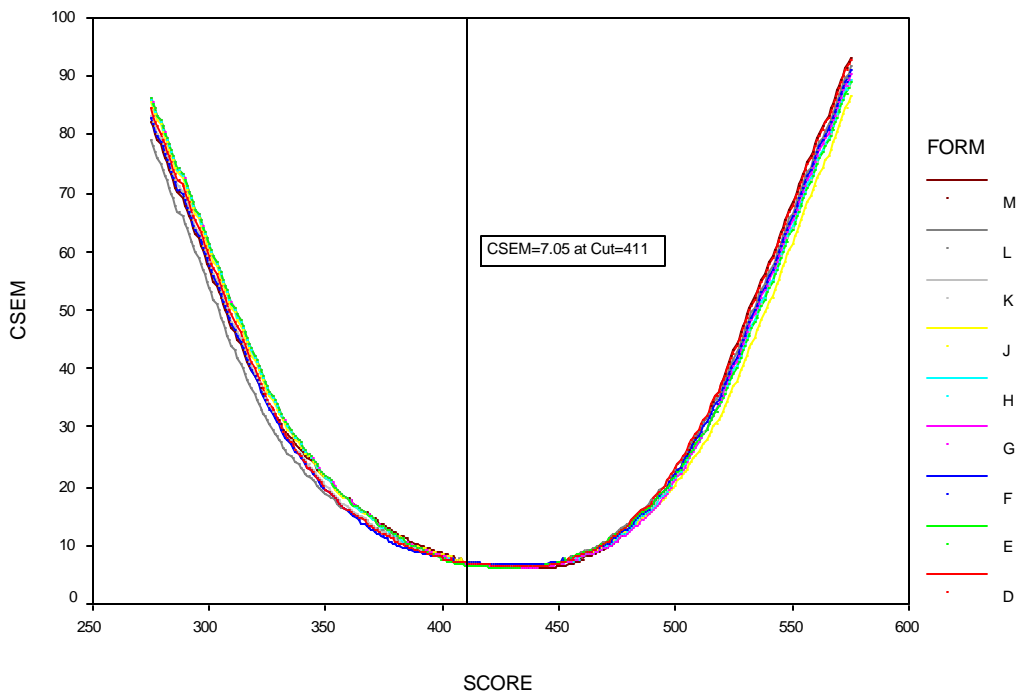
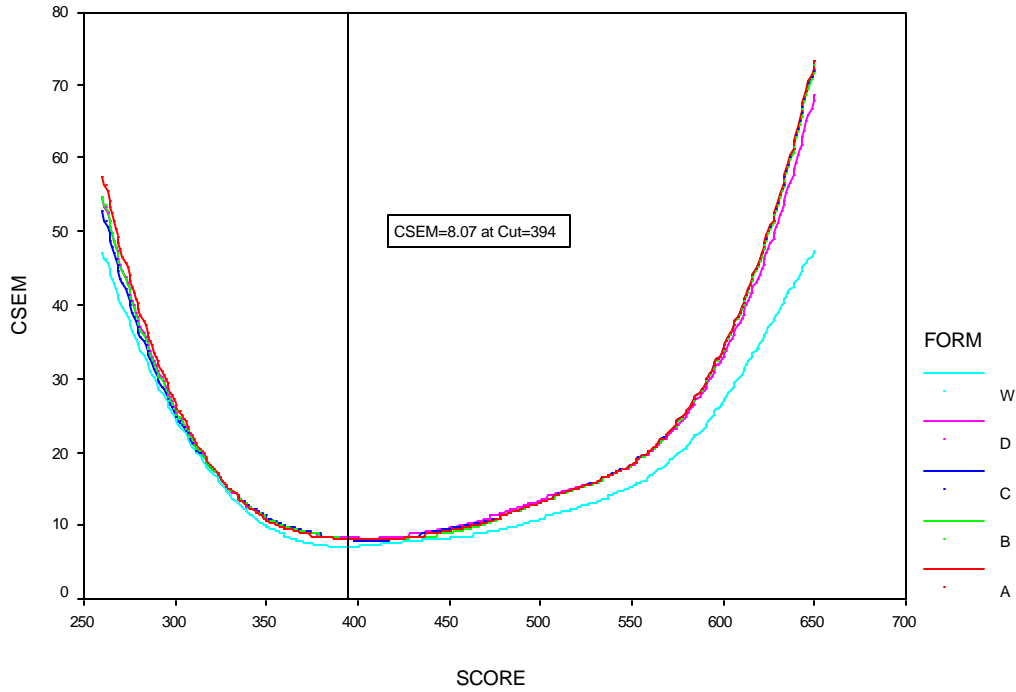


Figure 2d. Conditional Standard Errors for 2003 Geometry Assessments

Conditional Standard Errors

January 2003 Government



Conditional Standard Errors

May 2003 Government

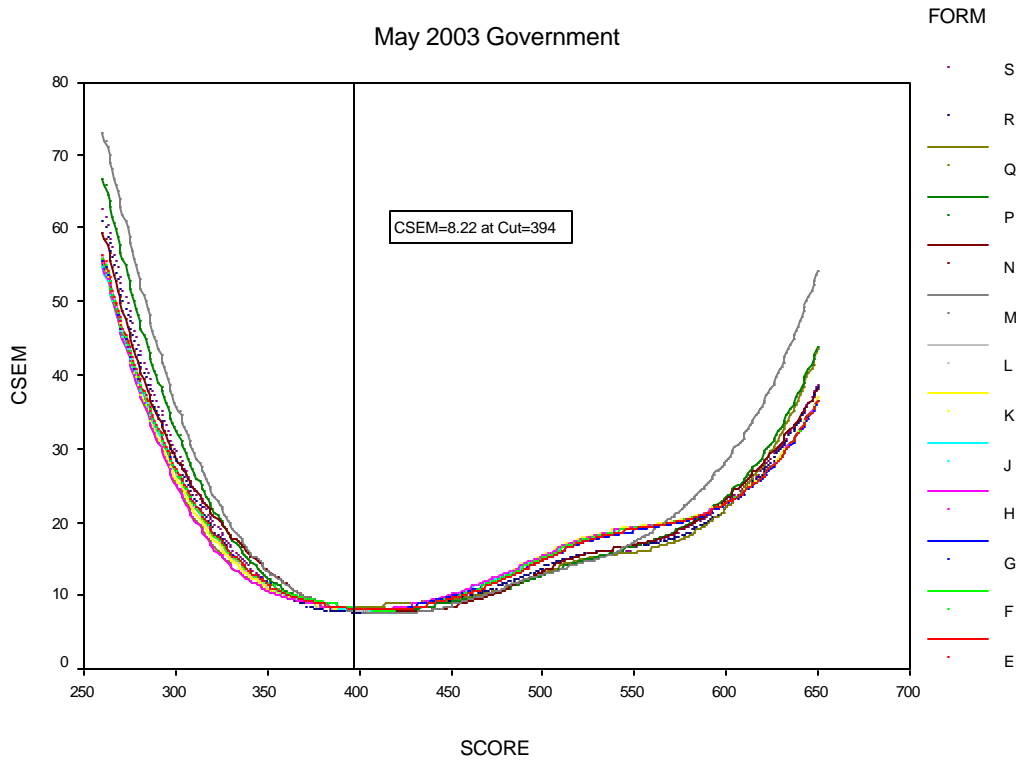


Figure 2e. Conditional Standard Errors for 2003 Government Assessments

References

- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of an EM algorithm. *Psychometrika*, 46, 443–459.
- Burket, G. R. (1988). ITEMSYS [Computer program]. Unpublished.
- Burket, G. R. (1991). PARDUX [Computer program]. Unpublished.
- Lord, F.M. (1980). Applications of Item Response Theory to Practical Testing Problems. Hillsdale, NJ: Erlbaum.
- Lord, F. M. and Novick, M.R. (1968). *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley.
- Phillips, S. E. (April, 2000). "Legal Corner: GI Forum v TEA." *NCME Newsletter* 8, 2, n.p.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 2, 125–145.
- Yen, W. M. (1991a). *How to evaluate FIT for Maryland*. Unpublished manuscript.
- Zwick, R., Donoghue, J.R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. *Journal of Educational Measurement*, 30, 233–251.

APPENDIX A

Item-level Statistics for Constructed Response Items

Table A1. Summary Statistics for January Constructed Response Items.

	Session	Item Number	(B)CR /(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*							Mean Raw Score	Mean Percent	Item-total Correlation**
						0	1	2	3	4	5	6			
Algebra															
Form A	1	6	B	2027	1866 (92)	18	29	30	23	.			1.59	53	0.43
	1	12	E	2027	1854 (91)	6	27	29	8	31			2.31	58	0.66
	1	17	B	2027	1647 (81)	6	56	29	9	.			1.40	47	0.55
	1	21	E	2027	1308 (65)	27	58	9	6	1			0.96	24	0.56
	2	30	B	2027	1552 (77)	15	63	17	4	.			1.11	37	0.57
	2	36	E	2027	1547 (76)	13	37	25	22	3			1.65	41	0.56
Form B	1	6	B	1787	1683 (94)	19	31	30	20	.			1.51	50	0.42
	1	12	E	1787	1647 (92)	4	21	32	8	35			2.50	63	0.65
	1	17	B	1787	1519 (85)	5	54	32	9	.			1.45	48	0.54
	1	21	E	1787	1188 (66)	24	56	13	6	1			1.04	26	0.56
	2	30	B	1787	1425 (80)	15	65	15	5	.			1.11	37	0.57
	2	36	E	1787	1399 (78)	11	35	26	25	4			1.77	44	0.51
Form W	1	6	B	1762	1610 (91)	11	59	11	19	.			1.37	46	0.47
	1	12	E	1762	1272 (72)	45	41	7	3	3			0.77	19	0.53
	1	17	B	1762	1451 (82)	9	71	9	11	.			1.22	41	0.55
	1	21	E	1762	1078 (61)	35	46	9	6	3			0.97	24	0.54
	2	30	B	1762	1234 (70)	37	42	16	4	.			0.88	29	0.57
	2	36	E	1762	1325 (75)	16	39	24	9	12			1.60	40	0.62
	2	41	B	1762	1352 (77)	14	25	32	30	.			1.77	59	0.39
	2	45	E	1762	1422 (81)	19	40	20	18	3			1.45	36	0.61
Biology															
Form A	1	5	B	2239	2091 (93)	9	49	38	4	<1			1.38	34	0.56
	1	12	B	2239	2105 (94)	3	37	51	9	1			1.68	42	0.58
	1	16	B	2239	2028 (91)	13	56	22	7	2			1.28	32	0.56
	1	23	B	2239	1928 (86)	7	51	36	5	1			1.43	36	0.68
	1	30	B	2239	1300 (58)	39	43	13	3	1			0.85	21	0.64

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

*** Although makeup forms are generally not included in our statistical analyses, the Geometry makeup forms were included in these analyses for reasons discussed previously in this report.

Table A1. Summary Statistics for January Constructed Response Items (cont.)

	Session	Item Number	(B)CR / (E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item-total Correlation**
						0	1	2	3	4	5			
Biology														
Form A	2	42	B	2239	1975 (88)	9	58	26	6	1		1.31	33	0.63
	2	50	B	2239	1631 (73)	34	32	20	12	3		1.17	29	0.75
Form B	1	5	B	1938	1816 (94)	8	46	42	5	<1		1.45	36	0.57
	1	12	B	1938	1834 (95)	3	34	53	9	1		1.71	43	0.58
	1	16	B	1938	1758 (91)	13	55	22	9	1		1.30	33	0.57
	1	23	B	1938	1674 (86)	5	49	37	8	1		1.50	38	0.67
	1	30	B	1938	1133 (58)	38	42	15	4	2		0.90	23	0.64
	2	42	B	1938	1730 (89)	8	57	27	6	1		1.35	34	0.59
	2	50	B	1938	1399 (72)	30	32	21	15	3		1.29	32	0.75
Form C	1	5	B	1890	1773 (94)	7	46	40	7	<1		1.47	37	0.59
	1	12	B	1890	1772 (94)	3	32	54	10	1		1.74	43	0.56
	1	16	B	1890	1713 (91)	13	54	24	7	2		1.31	33	0.59
	1	23	B	1890	1622 (86)	6	47	39	7	1		1.51	38	0.66
	1	30	B	1890	1132 (60)	37	45	13	3	1		0.87	22	0.63
	2	42	B	1890	1676 (89)	7	51	33	7	2		1.45	36	0.60
	2	50	B	1890	1372 (73)	29	32	21	14	4		1.32	33	0.74
Form D	1	5	B	1819	1738 (96)	7	47	40	6	<1		1.46	36	0.56
	1	12	B	1819	1751 (96)	3	33	54	9	1		1.72	43	0.55
	1	16	B	1819	1692 (93)	11	57	22	8	1		1.32	33	0.58
	1	23	B	1819	1599 (88)	5	48	39	8	1		1.52	38	0.68
	1	30	B	1819	1094 (60)	35	45	15	5	1		0.94	23	0.59
	2	42	B	1819	1638 (90)	6	55	29	7	2		1.42	36	0.64
	2	50	B	1819	1317 (72)	32	29	20	15	4		1.30	32	0.75
Form W	1	5	B	1819	1745 (96)	2	41	48	9	<1		1.65	41	0.50
	1	11	B	1819	1686 (93)	5	59	31	5	<1		1.35	34	0.54
	1	17	B	1819	1441 (79)	16	62	19	3	<1		1.09	27	0.64
	1	24	B	1819	1316 (72)	21	65	12	2	1		0.97	24	0.67

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

*** Although makeup forms are generally not included in our statistical analyses, the Geometry makeup forms were included in these analyses for reasons discussed previously in this report.

Table A1. Summary Statistics for January Constructed Response Items (cont.)

	Session	Item Number	(B)CR / (E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*							Mean Raw Score	Mean Percent	Item-total Correlation**
						0	1	2	3	4	5	6			
Biology															
Form W	1	32	B	1819	1352 (74)	26	56	15	3	<1			0.96	24	0.68
	2	42	B	1819	1240 (68)	32	54	10	3	1			0.86	21	0.62
	2	50	B	1819	1561 (86)	7	79	13	1	<1			1.08	27	0.50
	2	57	B	1819	1421 (78)	19	61	16	3	<1			1.04	26	0.61
	2	64	B	1819	1523 (84)	4	80	13	2	<1			1.14	28	0.46
English															
Form A	1	10	B	2370	2251 (95)	1	9	49	37	4	-	-	2.34	59	0.61
	1	27	B	2370	2058 (87)	4	13	46	32	4	-	-	2.19	55	0.68
	2	43	E	2370	2154 (91)	<1	5	19	27	34	13	2	3.37	56	0.64
Form B	1	10	B	2090	2021 (97)	1	7	49	40	4	-	-	2.40	60	0.55
	1	27	B	2090	1871 (90)	3	13	44	36	4	-	-	2.27	57	0.64
	2	43	E	2090	1930 (92)	<1	3	15	28	42	12	1	3.47	58	0.55
Form C	1	10	B	2019	1957 (97)	1	6	54	37	3	-	-	2.36	59	0.55
	1	27	B	2019	1792 (89)	3	12	50	32	3	-	-	2.20	55	0.61
	2	43	E	2019	1873 (93)	<1	3	15	28	42	12	1	3.46	58	0.59
Form W	1	8	B	1986	1905 (96)	2	13	53	28	3	-	-	2.17	54	0.53
	1	16	B	1986	1828 (92)	3	12	56	27	2	-	-	2.14	54	0.60
	2	47	E	1986	1857 (94)	<1	2	12	30	47	9	<1	3.49	58	0.52
	2	63	B	1986	1761 (89)	6	13	47	31	4	-	-	2.14	54	0.61
Geometry															
Form A	1	6	B	738	581 (79)	33	9	21	36	.			1.60	53	0.70
	1	12	E	738	616 (83)	42	42	4	6	6			0.93	23	0.61
	1	17	B	738	617 (84)	8	17	35	40	.			2.06	69	0.60
	1	21	E	738	557 (75)	24	53	16	5	2			1.09	27	0.66
	2	31	E	738	528 (72)	12	53	14	12	9			1.54	39	0.52
Form B	1	6	B	660	504 (76)	32	7	23	38	.			1.67	56	0.73
	1	12	E	660	539 (82)	42	40	6	5	6			0.93	23	0.64

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

*** Although makeup forms are generally not included in our statistical analyses, the Geometry makeup forms were included in these analyses for reasons discussed previously in this report.

Table A1. Summary Statistics for January Constructed Response Items (cont.)

	Session	Item Number	(B)CR /(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item-total Correlation**
						0	1	2	3	4	5			
Geometry														
Form B	1	17	B	660	553 (84)	8	18	38	36	.		2.03	68	0.54
	1	21	E	660	477 (72)	21	50	22	3	4		1.19	30	0.59
	2	31	E	660	461 (70)	13	52	15	11	9		1.52	38	0.55
Form C	1	6	B	640	501 (78)	32	6	22	41	.		1.71	57	0.67
	1	12	E	640	539 (84)	41	45	4	4	6		0.91	23	0.60
	1	17	B	640	559 (87)	10	19	36	35	.		1.95	65	0.56
	1	21	E	640	490 (77)	20	50	20	6	4		1.23	31	0.69
	2	31	E	640	467 (73)	12	55	11	12	10		1.54	39	0.50
Form W	1	6	B	629	578 (92)	6	48	36	11	.		1.52	51	0.53
	1	12	E	629	536 (85)	24	54	15	3	5		1.10	27	0.66
	1	17	B	629	454 (72)	22	48	21	9	.		1.17	39	0.57
	1	21	E	629	494 (79)	31	19	15	8	26		1.79	45	0.78
	2	31	E	629	430 (68)	70	19	7	2	2		0.47	12	0.60
	2	40	B	629	531 (84)	54	24	21	1	.		0.67	22	0.66
	2	44	E	629	466 (74)	29	42	20	8	2		1.11	28	0.68
Form X***	1	6	B	4861	3903 (80)	28	9	22	41	.		1.75	58	0.67
	1	12	E	4861	4092 (84)	45	38	4	4	8		0.92	23	0.60
	1	17	B	4861	4188 (86)	8	18	40	34	.		2.00	67	0.54
	1	21	E	4861	3594 (74)	21	48	22	7	3		1.24	31	0.64
	2	31	E	4861	3663 (75)	8	54	12	14	12		1.70	42	0.59
Form Y***	1	6	B	1217	839 (69)	42	9	19	30	.		1.36	45	0.68
	1	12	E	1217	916 (75)	54	33	4	4	5		0.72	18	0.58
	1	17	B	1217	938 (77)	12	25	38	25	.		1.77	59	0.51
	1	21	E	1217	761 (63)	27	51	17	3	2		1.01	25	0.60
	2	31	E	1217	799 (66)	15	56	12	11	6		1.35	34	0.57

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

*** Although makeup forms are generally not included in our statistical analyses, the Geometry makeup forms were included in these analyses for reasons discussed previously in this report.

Table A1. Summary Statistics for January Constructed Response Items (cont.)

	Session	Item Number	(B)CR/(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*							Mean Raw Score	Mean Percent	Item-total Correlation**
						0	1	2	3	4	5	6			
Government															
Form A	1	6	B	2380	2273 (96)	14	23	52	12	.			1.61	40	0.56
	1	14	B	2380	2163 (91)	3	19	64	14	0			1.90	48	0.49
	1	20	B	2380	2084 (88)	6	21	58	14	1			1.82	46	0.61
	1	28	E	2380	2034 (85)	7	51	36	5	0			1.40	35	0.64
	1	34	B	2380	2009 (84)	7	27	53	12	0			1.70	43	0.67
	2	46	B	2380	2177 (91)	2	26	56	15	1			1.86	46	0.57
	2	52	B	2380	2076 (87)	4	29	53	14	1			1.79	45	0.66
	2	57	B	2380	1942 (82)	14	36	40	10	0			1.46	36	0.66
Form B	1	6	B	2038	1965 (96)	11	19	54	16	.			1.74	44	0.52
	1	14	B	2038	1866 (92)	2	17	63	17	1			1.97	49	0.50
	1	20	B	2038	1775 (87)	5	19	59	16	0			1.88	47	0.58
	1	28	E	2038	1759 (86)	9	55	32	4	0			1.30	33	0.66
	1	34	B	2038	1762 (86)	5	24	54	16	0			1.82	46	0.65
	2	46	B	2038	1863 (91)	1	22	59	17	1			1.94	49	0.57
	2	52	B	2038	1786 (88)	2	23	60	14	1			1.88	47	0.63
	2	57	B	2038	1669 (82)	10	34	45	10	0			1.56	39	0.64
Form C	1	6	B	2012	1946 (97)	10	19	53	18	0			1.78	45	0.58
	1	14	B	2012	1842 (92)	2	16	60	22	1			2.03	51	0.52
	1	20	B	2012	1751 (87)	6	18	57	18	1			1.89	47	0.60
	1	28	E	2012	1738 (86)	9	58	29	4	.			1.28	32	0.66
	1	34	B	2012	1721 (86)	7	24	54	15	0			1.78	44	0.64
	2	46	B	2012	1738 (86)	12	34	44	10	0			1.53	38	0.61
	2	52	B	2012	1768 (88)	2	27	56	15	0			1.84	46	0.64
	2	57	B	2012	1803 (90)	2	32	53	14	0			1.78	45	0.58

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

*** Although makeup forms are generally not included in our statistical analyses, the Geometry makeup forms were included in these analyses for reasons discussed previously in this report.

Table A1. Summary Statistics for January Constructed Response Items (cont.)

	Session	Item Number	(B)CR / (E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*							Mean Raw Score	Mean Percent	Item-total Correlation**
						0	1	2	3	4	5	6			
Government															
Form D	1	6	B	1956	1884 (96)	11	21	57	11	0		1.69	42	0.54	
	1	14	B	1956	1797 (92)	2	17	65	16	1		1.97	49	0.53	
	1	20	B	1956	1739 (89)	5	17	60	16	1		1.91	48	0.60	
	1	28	E	1956	1688 (86)	7	54	33	5	0		1.36	34	0.68	
	1	34	B	1956	1712 (88)	4	26	55	13	1		1.80	45	0.63	
	2	46	B	1956	1811 (93)	2	29	54	14	1		1.81	45	0.57	
	2	52	B	1956	1740 (89)	3	25	58	13	1		1.83	46	0.63	
	2	57	B	1956	1627 (83)	12	35	43	9	0		1.52	38	0.62	
Form W	1	6	B	1917	1736 (91)	10	32	51	6	0		1.54	38	0.60	
	1	14	B	1917	1847 (96)	4	23	60	13	0		1.82	46	0.57	
	1	20	B	1917	1649 (86)	10	24	54	12	0		1.68	42	0.67	
	1	28	E	1917	1494 (78)	17	40	37	7	.		1.34	34	0.69	
	1	34	B	1917	1652 (86)	8	35	48	9	0		1.58	39	0.63	
	2	46	B	1917	1440 (75)	18	36	38	7	0		1.37	34	0.66	
	2	52	B	1917	1566 (82)	21	40	35	4	0		1.24	31	0.60	
	2	58	B	1917	1565 (82)	11	35	48	5	0		1.49	37	0.64	
	2	71	B	1917	1576 (82)	11	29	51	8	0		1.57	39	0.64	

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

*** Although makeup forms are generally not included in our statistical analyses, the Geometry makeup forms were included in these analyses for reasons discussed previously in this report.

Table A2. Summary Statistics for May Constructed Response Items

	Session	Item Number	(B)CR /(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Algebra														
Form C	1	6	B	6585	6117 (93)	20	28	28	24	-		1.55	52	0.47
	1	12	E	6585	6006 (91)	5	22	24	7	42		2.59	65	0.74
	1	17	B	6585	5597 (85)	6	50	30	15	-		1.52	51	0.63
	1	21	E	6585	4810 (73)	22	51	13	10	4		1.22	31	0.58
	2	30	B	6585	5279 (80)	13	53	22	11	-		1.31	44	0.62
	2	36	E	6585	5242 (80)	12	30	23	27	8		1.90	47	0.60
Form D	1	6	B	5639	5346 (95)	18	28	29	25	-		1.62	54	0.41
	1	12	E	5639	5253 (93)	3	19	25	6	46		2.73	68	0.70
	1	17	B	5639	4967 (88)	5	47	32	16	-		1.60	53	0.60
	1	21	E	5639	4220 (75)	19	54	12	10	5		1.28	32	0.55
	2	30	B	5639	4623 (82)	11	52	24	13	-		1.38	46	0.62
	2	36	E	5639	4620 (82)	8	30	23	30	9		2.02	50	0.53
Form E	1	6	B	5559	5286 (95)	17	29	29	24	-		1.60	53	0.41
	1	12	E	5559	5221 (94)	3	20	23	7	47		2.74	69	0.71
	1	17	B	5559	4926 (89)	5	47	32	17	-		1.60	53	0.61
	1	21	E	5559	4175 (75)	19	54	13	9	5		1.27	32	0.55
	2	30	B	5559	4605 (83)	11	55	23	11	-		1.35	45	0.60
	2	36	E	5559	4581 (82)	9	30	22	29	9		1.98	49	0.58
Form F	1	6	B	5468	5185 (95)	17	29	29	25	-		1.63	54	0.40
	1	12	E	5468	5100 (93)	3	19	25	6	47		2.76	69	0.70
	1	17	B	5468	4847 (89)	5	47	33	16	-		1.60	53	0.61
	1	21	E	5468	4076 (75)	18	57	11	9	5		1.25	31	0.54
	2	30	B	5468	4509 (82)	10	54	24	12	-		1.38	46	0.61
	2	36	E	5468	4517 (83)	8	31	22	31	9		2.02	50	0.55
Form G	1	6	B	5420	5139 (95)	18	28	30	24	-		1.60	53	0.41
	1	12	E	5420	5048 (93)	3	20	25	7	44		2.69	67	0.68
	1	17	B	5420	4757 (88)	4	48	32	16	-		1.60	53	0.60

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR /(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Algebra														
Form G	1	21	E	5420	4018 (74)	19	53	14	9	4		1.25	31	0.56
	2	30	B	5420	4485 (83)	12	53	23	12	-		1.35	45	0.62
	2	36	E	5420	4402 (81)	9	31	22	29	9		1.96	49	0.54
Form H	1	6	B	5405	5110 (95)	17	27	30	26	-		1.66	55	0.41
	1	12	E	5405	5036 (93)	4	20	24	7	46		2.72	68	0.71
	1	17	B	5405	4783 (88)	4	48	33	15	-		1.59	53	0.60
	1	21	E	5405	4059 (75)	18	55	12	9	5		1.28	32	0.54
	2	30	B	5405	4446 (82)	12	53	22	12	-		1.35	45	0.61
	2	36	E	5405	4451 (82)	8	30	24	31	8		2.01	50	0.56
Form J	1	6	B	5340	5073 (95)	17	29	29	25	-		1.63	54	0.39
	1	12	E	5340	5020 (94)	4	19	25	7	46		2.73	68	0.70
	1	17	B	5340	4761 (89)	4	47	33	16	-		1.61	54	0.61
	1	21	E	5340	4103 (77)	19	54	12	10	5		1.28	32	0.56
	2	30	B	5340	4408 (83)	12	54	23	11	-		1.34	45	0.62
	2	36	E	5340	4421 (83)	9	31	22	29	9		1.98	50	0.55
Form K	1	6	B	5256	4979 (95)	16	30	29	25	-		1.64	55	0.40
	1	12	E	5256	4911 (93)	3	18	23	7	48		2.79	70	0.70
	1	17	B	5256	4637 (88)	5	47	33	16	-		1.60	53	0.60
	1	21	E	5256	3936 (75)	18	55	12	9	5		1.27	32	0.55
	2	30	B	5256	4349 (83)	12	53	24	12	-		1.37	46	0.61
	2	36	E	5256	4347 (83)	8	30	23	31	8		1.99	50	0.55
Form L	1	6	B	5194	4924 (95)	16	29	31	25	-		1.64	55	0.42
	1	12	E	5194	4851 (93)	3	19	25	7	47		2.76	69	0.70
	1	17	B	5194	4539 (87)	4	47	33	16	-		1.61	54	0.58
	1	21	E	5194	3867 (74)	20	55	11	9	4		1.24	31	0.53
	2	30	B	5194	4301 (83)	12	53	23	12	-		1.36	45	0.60
	2	36	E	5194	4258 (82)	8	30	22	32	9		2.04	51	0.54

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR/(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Algebra														
Form M	1	6	B	5134	4862 (95)	4	19	24	53	-		2.26	75	0.53
	1	15	B	5134	4372 (85)	61	21	13	5	-		0.62	21	0.47
	1	20	E	5134	3952 (77)	35	19	36	5	6		1.27	32	0.68
	2	31	B	5134	3795 (74)	27	35	22	16	-		1.28	43	0.65
	2	36	E	5134	4466 (87)	11	35	42	4	8		1.65	41	0.53
	2	40	B	5134	3513 (68)	72	24	4	<1	-		0.31	10	0.40
Form N	1	6	B	5069	4270 (84)	35	37	18	9	-		1.02	34	0.62
	1	12	E	5069	4448 (88)	9	30	24	20	17		2.08	52	0.72
	1	17	B	5069	4496 (89)	25	25	43	7	-		1.33	44	0.57
	2	36	E	5069	4043 (80)	8	42	19	20	11		1.82	46	0.67
	2	41	B	5069	4235 (84)	8	27	28	37	-		1.95	65	0.57
	2	45	E	5069	4272 (84)	6	15	20	18	41		2.75	69	0.67
Biology														
Form E	1	5	B	4965	4594 (93)	9	35	49	7	<1		1.54	38	0.56
	1	12	B	4965	4622 (93)	3	29	57	10	1		1.77	44	0.60
	1	16	B	4965	4359 (88)	14	55	21	8	3		1.31	33	0.62
	1	23	B	4965	4223 (85)	6	47	38	7	1		1.51	38	0.67
	1	30	B	4965	2988 (60)	35	48	12	3	1		0.86	21	0.62
	2	42	B	4965	4358 (88)	8	57	28	6	1		1.36	34	0.63
	2	50	B	4965	3549 (71)	33	26	23	14	4		1.31	33	0.76
Form F	1	5	B	4147	3904 (94)	7	35	52	6	<1		1.59	40	0.54
	1	12	B	4147	3922 (95)	2	25	61	11	1		1.84	46	0.55
	1	16	B	4147	3715 (90)	12	55	22	9	3		1.38	34	0.63
	1	23	B	4147	3606 (87)	4	46	39	9	2		1.59	40	0.66
	1	30	B	4147	2519 (61)	33	46	16	4	1		0.95	24	0.62
	2	42	B	4147	3722 (90)	7	56	29	6	2		1.40	35	0.60
	2	50	B	4147	3016 (73)	29	26	22	18	5		1.44	36	0.73

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR / (E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Biology														
Form G	1	5	B	4126	3877 (94)	6	36	51	7	<1		1.60	40	0.53
	1	12	B	4126	3902 (95)	2	26	61	10	1		1.81	45	0.56
	1	16	B	4126	3690 (89)	12	53	23	9	3		1.38	34	0.63
	1	23	B	4126	3537 (86)	4	48	39	8	2		1.56	39	0.65
	1	30	B	4126	2517 (61)	31	47	16	4	2		0.98	25	0.62
	2	42	B	4126	3660 (89)	7	53	30	8	2		1.45	36	0.59
	2	50	B	4126	2991 (72)	26	29	23	17	6		1.47	37	0.74
Form H	1	5	B	4099	3846 (94)	6	36	50	8	<1		1.60	40	0.53
	1	12	B	4099	3868 (94)	1	26	59	12	1		1.86	46	0.55
	1	16	B	4099	3665 (89)	12	52	24	10	2		1.38	35	0.62
	1	23	B	4099	3582 (87)	4	47	39	9	1		1.56	39	0.64
	1	30	B	4099	2500 (61)	31	51	13	4	1		0.93	23	0.62
	2	42	B	4099	3647 (89)	6	55	31	6	2		1.42	36	0.60
	2	50	B	4099	2984 (73)	30	26	22	17	5		1.42	36	0.73
Form J	1	5	B	4123	3870 (94)	7	36	49	7	<1		1.57	39	0.55
	1	12	B	4123	3901 (95)	2	27	61	10	1		1.81	45	0.54
	1	16	B	4123	3725 (90)	12	54	21	10	3		1.36	34	0.60
	1	23	B	4123	3598 (87)	4	50	37	8	1		1.52	38	0.65
	1	30	B	4123	2554 (62)	31	44	18	5	2		1.02	25	0.62
	2	42	B	4123	3735 (91)	6	52	32	8	2		1.48	37	0.61
	2	50	B	4123	3061 (74)	27	27	22	17	6		1.47	37	0.75
Form K	1	5	B	4051	3814 (94)	6	33	52	9	<1		1.65	41	0.54
	1	12	B	4051	3807 (94)	1	26	59	13	1		1.87	47	0.54
	1	16	B	4051	3602 (89)	11	54	22	10	3		1.40	35	0.62
	1	23	B	4051	3485 (86)	5	44	40	9	2		1.59	40	0.64
	1	30	B	4051	2491 (61)	31	51	14	3	1		0.92	23	0.61
	2	42	B	4051	3617 (89)	7	57	29	5	2		1.39	35	0.60

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR /(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*							Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5	6			
Biology															
Form K	2	50	B	4051	2963 (73)	30	24	25	15	5			1.41	35	0.72
Form L	1	5	B	4017	3759 (94)	6	36	51	6	<1			1.58	39	0.50
	1	12	B	4017	3807 (95)	2	26	61	10	1			1.83	46	0.53
	1	16	B	4017	3601 (90)	12	56	22	9	3			1.35	34	0.59
	1	23	B	4017	3476 (87)	4	50	38	7	1			1.53	38	0.61
	1	30	B	4017	2544 (63)	32	47	16	4	1			0.97	24	0.62
	2	42	B	4017	3585 (89)	6	55	29	8	2			1.44	36	0.61
	2	50	B	4017	2919 (73)	29	27	23	16	5			1.41	35	0.73
Form M	1	7	B	4004	3616 (90)	8	70	20	1	<1			1.15	29	0.54
	1	36	B	4004	3213 (80)	14	67	17	2	<1			1.06	27	0.59
	2	53	B	4004	3387 (85)	5	50	33	9	2			1.53	38	0.64
	2	57	B	4004	3455 (86)	4	47	35	10	4			1.63	41	0.62
	2	63	B	4004	2981 (74)	3	67	23	7	1			1.36	34	0.58
	2	69	B	4004	3063 (76)	4	53	30	11	2			1.56	39	0.68
Form N	1	6	B	3970	3749 (94)	2	53	37	6	1			1.51	38	0.55
	1	15	B	3970	3666 (92)	1	52	37	8	1			1.57	39	0.59
	1	22	B	3970	2938 (74)	14	66	17	2	<1			1.08	27	0.52
	1	30	B	3970	2992 (75)	21	57	16	6	1			1.09	27	0.68
	2	38	B	3970	3474 (88)	13	35	21	24	7			1.77	44	0.71
	2	55	B	3970	3415 (86)	9	37	40	12	2			1.61	40	0.69
Form P	1	6	B	3910	3683 (94)	2	33	41	19	4			1.91	48	0.64
	1	14	B	3910	2476 (63)	19	38	28	12	4			1.44	36	0.72
	1	20	B	3910	3165 (81)	10	57	21	8	4			1.39	35	0.73
	1	26	B	3910	3101 (79)	16	51	20	9	4			1.34	33	0.69
	1	30	B	3910	3294 (84)	2	49	33	11	4			1.65	41	0.70
	2	38	B	3910	3461 (89)	13	69	18	<1	<1			1.06	26	0.49
	2	45	B	3910	3360 (86)	12	44	28	13	3			1.52	38	0.67

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR /(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*							Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5	6			
Biology															
Form P	2	59	B	3910	3266 (84)	7	37	40	13	3			1.68	42	0.70
Form Q	1	6	B	3866	3623 (94)	3	65	27	4	1			1.35	34	0.55
	1	10	B	3866	2805 (73)	23	50	20	6	1			1.11	28	0.71
	1	17	B	3866	3487 (90)	9	47	18	20	6			1.68	42	0.67
	1	23	B	3866	3372 (87)	11	52	26	10	1			1.38	34	0.70
	1	27	B	3866	2591 (67)	30	47	18	4	1			0.99	25	0.73
	2	42	B	3866	3311 (86)	6	61	25	8	<1			1.34	34	0.57
	2	47	B	3866	3163 (82)	12	51	29	8	<1			1.33	33	0.69
Form R	1	6	B	3821	3548 (93)	3	58	31	6	1			1.43	36	0.59
	1	14	B	3821	3389 (89)	2	43	42	12	2			1.69	42	0.63
	1	19	B	3821	3368 (88)	9	68	17	5	1			1.22	31	0.66
	1	24	B	3821	2920 (76)	9	59	25	5	1			1.30	32	0.66
	2	42	B	3821	3052 (80)	41	37	11	9	3			0.97	24	0.65
	2	48	B	3821	3167 (83)	10	51	29	8	1			1.38	34	0.71
	2	61	B	3821	3262 (85)	4	43	42	9	2			1.62	40	0.67
English															
Form D	1	10	B	5831	5522 (95)	2	12	53	30	3	-	-	2.20	55	0.63
	1	27	B	5831	5053 (87)	5	19	46	27	4	-	-	2.06	51	0.67
	2	43	E	5831	5220 (90)	<1	4	18	26	36	14	2	3.44	57	0.61
Form E	1	10	B	4797	4646 (97)	1	9	55	32	3	-	-	2.27	57	0.57
	1	27	B	4797	4312 (90)	3	16	47	30	4	-	-	2.15	54	0.63
	2	43	E	4797	4456 (93)	<1	2	14	27	40	15	2	3.57	60	0.55
Form F	1	10	B	4806	4642 (97)	1	8	56	32	3	-	-	2.29	57	0.54
	1	27	B	4806	4334 (90)	4	14	49	30	4	-	-	2.15	54	0.62
	2	43	E	4806	4433 (92)	<1	2	13	26	40	15	2	3.60	60	0.53
Form G	1	10	B	4772	4623 (97)	1	8	56	33	2	-	-	2.28	57	0.54
	1	27	B	4772	4290 (90)	3	15	49	30	3	-	-	2.14	53	0.62

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR/(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*							Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5	6			
English															
Form G	2	43	E	4772	4422 (93)	<1	2	15	27	40	15	2	3.56	59	0.52
Form H	1	10	B	4775	4620 (97)	1	8	56	33	3	-	-	2.28	57	0.57
	1	27	B	4775	4298 (90)	3	15	48	30	4	-	-	2.17	54	0.63
	2	43	E	4775	4445 (93)	<1	2	14	25	42	14	3	3.59	60	0.55
Form J	1	10	B	4720	4563 (97)	1	8	55	33	3	-	-	2.27	57	0.55
	1	27	B	4720	4253 (90)	3	15	45	32	4	-	-	2.18	55	0.63
	2	43	E	4720	4381 (93)	<1	3	14	25	40	16	3	3.61	60	0.54
Form K	1	10	B	4673	4544 (97)	1	8	52	36	3	-	-	2.33	58	0.55
	1	27	B	4673	4190 (90)	3	14	44	35	5	-	-	2.25	56	0.62
	2	43	E	4673	4329 (93)	<1	2	14	26	39	16	3	3.60	60	0.53
Form L	1	10	B	4600	4433 (96)	1	8	51	37	3	-	-	2.35	59	0.57
	1	27	B	4600	4118 (90)	3	15	44	33	5	-	-	2.22	56	0.65
	2	43	E	4600	4199 (91)	<1	2	15	25	40	16	3	3.61	60	0.56
Form M	1	3	B	4596	4410 (96)	3	15	48	32	3	-	-	2.16	54	0.50
	1	10	B	4596	4244 (92)	3	19	49	26	2	-	-	2.04	51	0.64
	2	54	E	4596	4196 (91)	<1	4	17	27	39	11	2	3.44	57	0.55
Form N	1	5	B	4508	4245 (94)	3	26	47	22	2	-	-	1.93	48	0.54
	1	27	B	4508	4084 (91)	1	24	55	19	1	-	-	1.96	49	0.60
	2	43	E	4508	4223 (94)	<1	2	15	30	39	11	2	3.47	58	0.55
	2	51	B	4508	3914 (87)	4	32	45	18	2	-	-	1.81	45	0.61
Form P	1	10	B	4483	4198 (94)	4	25	49	20	2	-	-	1.93	48	0.60
	1	27	B	4483	4055 (90)	4	29	45	19	2	-	-	1.86	46	0.62
	2	43	E	4483	4020 (90)	<1	3	14	29	41	11	2	3.50	58	0.58
	2	58	B	4483	3895 (87)	3	15	53	26	3	-	-	2.10	52	0.62
Geometry															
Form D	1	6	B	5790	4563 (79)	28	10	21	41	-			1.75	58	0.72
	1	12	E	5790	4803 (83)	45	36	5	4	9			0.97	24	0.64

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR / (E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Geometry														
Form D	1	17	B	5790	4960 (86)	9	16	37	38	-		2.03	68	0.60
	1	21	E	5790	4266 (74)	20	46	21	8	5		1.31	33	0.67
	2	31	E	5790	4342 (75)	15	50	15	8	11		1.49	37	0.62
Form E	1	6	B	5069	4028 (79)	25	9	22	45	-		1.86	62	0.72
	1	12	E	5069	4242 (84)	42	38	6	4	10		1.02	25	0.63
	1	17	B	5069	4361 (86)	8	15	37	40	-		2.10	70	0.56
	1	21	E	5069	3777 (75)	18	48	21	9	6		1.37	34	0.63
	2	31	E	5069	3825 (75)	13	51	16	8	12		1.55	39	0.61
Form F	1	6	B	4993	3986 (80)	24	9	21	46	-		1.90	63	0.69
	1	12	E	4993	4154 (83)	42	38	5	4	11		1.04	26	0.63
	1	17	B	4993	4279 (86)	7	15	37	41	-		2.12	71	0.58
	1	21	E	4993	3654 (73)	17	46	24	8	4		1.37	34	0.64
	2	31	E	4993	3766 (75)	13	52	16	8	12		1.53	38	0.61
Form G	1	6	B	4997	3965 (79)	23	9	22	45	-		1.90	63	0.68
	1	12	E	4997	4190 (84)	42	37	6	5	11		1.05	26	0.63
	1	17	B	4997	4283 (86)	8	15	38	39	-		2.09	70	0.53
	1	21	E	4997	3699 (74)	17	47	23	9	4		1.36	34	0.64
	2	31	E	4997	3819 (76)	12	52	16	8	12		1.56	39	0.61
Form H	1	6	B	4912	3936 (80)	24	8	23	45	-		1.89	63	0.68
	1	12	E	4912	4122 (84)	44	37	5	5	10		1.00	25	0.64
	1	17	B	4912	4228 (86)	7	15	37	41	-		2.12	71	0.56
	1	21	E	4912	3733 (76)	17	48	22	9	4		1.35	34	0.64
	2	31	E	4912	3740 (76)	11	52	16	8	12		1.57	39	0.59
Form J	1	6	B	4874	3918 (80)	24	9	21	45	-		1.87	62	0.69
	1	12	E	4874	4167 (85)	42	39	6	4	9		1.01	25	0.63
	1	17	B	4874	4276 (88)	6	15	37	41	-		2.14	71	0.56
	1	21	E	4874	3715 (76)	17	48	23	7	4		1.33	33	0.62

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR /(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Geometry														
Form J	2	31	E	4874	3722 (76)	12	52	15	9	13		1.59	40	0.62
Form K	1	6	B	4813	3892 (81)	24	9	22	45	-		1.88	63	0.71
	1	12	E	4813	4044 (84)	40	38	6	6	10		1.08	27	0.63
	1	17	B	4813	4108 (85)	7	16	37	40	-		2.10	70	0.57
	1	21	E	4813	3624 (75)	17	46	22	8	7		1.41	35	0.64
	2	31	E	4813	3677 (76)	11	53	15	9	12		1.58	40	0.60
Form L	1	6	B	4782	3822 (80)	24	8	21	46	-		1.90	63	0.67
	1	12	E	4782	3968 (83)	41	39	5	5	10		1.04	26	0.62
	1	17	B	4782	4089 (86)	8	14	38	40	-		2.11	70	0.55
	1	21	E	4782	3575 (75)	18	48	21	7	6		1.36	34	0.61
	2	31	E	4782	3641 (76)	12	55	14	8	11		1.53	38	0.58
Form M	1	12	B	4668	3470 (74)	13	22	40	26	-		1.78	59	0.65
	1	16	E	4668	3688 (79)	11	62	14	4	9		1.38	34	0.57
	1	22	E	4668	3567 (76)	27	19	25	10	18		1.74	43	0.76
	2	35	B	4668	3585 (77)	28	41	13	18	-		1.21	40	0.63
	2	40	E	4668	3640 (78)	15	31	29	9	16		1.81	45	0.75
Government														
Form E	1	6	B	4897	4606 (94)	12	18	57	13	<1		1.72	43	0.58
	1	14	B	4897	4359 (89)	3	13	68	15	1		1.98	50	0.53
	1	20	B	4897	4161 (85)	6	15	62	17	1		1.91	48	0.65
	1	28	E	4897	4027 (82)	7	53	32	8	<1		1.43	36	0.70
	1	34	B	4897	4142 (85)	8	20	59	13	1		1.79	45	0.66
	2	46	B	4897	4356 (89)	2	26	52	18	1		1.89	47	0.61
	2	52	B	4897	4172 (85)	4	26	49	20	1		1.88	47	0.67
	2	57	B	4897	3836 (78)	11	35	39	14	1		1.58	39	0.68
Form F	1	6	B	4005	3779 (94)	8	14	64	14	1		1.85	46	0.55
	1	14	B	4005	3579 (89)	2	9	72	17	1		2.05	51	0.50

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR/(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Government														
Form F	1	20	B	4005	3501 (87)	4	11	65	19	1		2.01	50	0.60
	1	28	E	4005	3383 (84)	3	48	39	10	<1		1.55	39	0.66
	1	34	B	4005	3437 (86)	4	17	63	16	1		1.92	48	0.63
	2	46	B	4005	3615 (90)	1	21	57	20	1		1.97	49	0.57
	2	52	B	4005	3471 (87)	2	22	55	20	1		1.96	49	0.63
	2	57	B	4005	3245 (81)	7	34	44	15	1		1.67	42	0.63
Form G	1	6	B	4037	3843 (95)	9	15	61	15	<1		1.84	46	0.55
	1	14	B	4037	3626 (90)	2	9	69	19	<1		2.08	52	0.48
	1	20	B	4037	3520 (87)	3	12	65	20	1		2.04	51	0.59
	1	28	E	4037	3444 (85)	4	50	37	9	<1		1.51	38	0.66
	1	34	B	4037	3498 (87)	3	18	61	17	1		1.94	48	0.61
	2	46	B	4037	3659 (91)	1	21	58	19	<1		1.97	49	0.56
	2	52	B	4037	3519 (87)	2	22	55	20	1		1.96	49	0.63
	2	57	B	4037	3344 (83)	8	34	44	15	1		1.67	42	0.65
Form H	1	6	B	3957	3750 (95)	9	14	62	15	<1		1.85	46	0.54
	1	14	B	3957	3552 (90)	2	8	69	20	1		2.10	52	0.48
	1	20	B	3957	3437 (87)	4	10	67	20	1		2.04	51	0.60
	1	28	E	3957	3339 (84)	4	50	37	9	<1		1.51	38	0.65
	1	34	B	3957	3372 (85)	4	15	63	17	1		1.94	49	0.62
	2	46	B	3957	3600 (91)	1	22	58	18	<1		1.95	49	0.55
	2	52	B	3957	3476 (88)	2	21	57	19	<1		1.95	49	0.62
	2	57	B	3957	3263 (82)	9	33	45	13	1		1.63	41	0.63
Form J	1	6	B	3971	3787 (95)	9	14	61	16	<1		1.85	46	0.54
	1	14	B	3971	3567 (90)	2	9	68	20	1		2.08	52	0.50
	1	20	B	3971	3464 (87)	4	11	63	22	1		2.06	52	0.61
	1	28	E	3971	3356 (85)	3	48	39	9	<1		1.55	39	0.68
	1	34	B	3971	3402 (86)	4	17	60	17	1		1.94	48	0.64

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR/(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Government														
Form J	2	46	B	3971	3587 (90)	1	21	59	18	1		1.96	49	0.56
	2	52	B	3971	3465 (87)	2	22	56	19	1		1.95	49	0.62
	2	57	B	3971	3285 (83)	8	34	44	13	1		1.64	41	0.64
Form K	1	6	B	3954	3784 (96)	8	14	60	18	<1		1.88	47	0.56
	1	14	B	3954	3567 (90)	2	9	66	22	1		2.10	52	0.50
	1	20	B	3954	3476 (88)	5	11	61	23	1		2.04	51	0.61
	1	28	E	3954	3340 (84)	5	50	35	10	<1		1.49	37	0.67
	1	34	B	3954	3427 (87)	4	16	59	20	1		1.97	49	0.64
	2	46	B	3954	3600 (91)	1	21	60	18	1		1.96	49	0.57
	2	52	B	3954	3466 (88)	3	21	57	18	1		1.94	48	0.64
	2	57	B	3954	3275 (83)	8	34	45	12	1		1.63	41	0.64
Form L	1	6	B	3977	3766 (95)	7	15	60	18	<1		1.89	47	0.52
	1	14	B	3977	3583 (90)	2	11	64	22	1		2.09	52	0.51
	1	20	B	3977	3510 (88)	3	11	61	24	1		2.08	52	0.60
	1	28	E	3977	3390 (85)	4	51	35	9	<1		1.50	38	0.66
	1	34	B	3977	3422 (86)	3	18	59	19	1		1.96	49	0.63
	2	46	B	3977	3611 (91)	1	21	61	17	<1		1.95	49	0.56
	2	52	B	3977	3510 (88)	2	21	59	18	1		1.94	48	0.62
	2	57	B	3977	3303 (83)	6	34	48	12	<1		1.67	42	0.63
Form M	1	6	B	3900	3631 (93)	6	32	45	15	1		1.74	43	0.67
	1	14	B	3900	3438 (88)	10	35	37	16	2		1.64	41	0.74
	1	20	B	3900	3331 (85)	13	26	39	21	2		1.72	43	0.76
	1	28	E	3900	3174 (81)	3	17	57	23	<1		2.00	50	0.59
	1	34	B	3900	3038 (78)	12	37	39	11	1		1.51	38	0.71
	2	46	B	3900	3282 (84)	5	38	45	12	<1		1.66	41	0.64
	2	52	B	3900	3229 (83)	10	41	38	11	<1		1.50	38	0.65

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR/(E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*							Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5	6			
Government															
Form N	1	6	B	3849	3586 (93)	4	25	53	19	1			1.88	47	0.60
	1	14	B	3849	3441 (89)	8	27	49	16	1			1.75	44	0.67
	1	20	B	3849	3225 (84)	12	26	40	21	1			1.73	43	0.73
	1	28	E	3849	3069 (80)	2	28	49	21	<1			1.89	47	0.63
	1	34	B	3849	3279 (85)	5	31	47	17	<1			1.77	44	0.66
	2	46	B	3849	3212 (83)	29	33	29	9	<1			1.18	30	0.59
	2	52	B	3849	3368 (88)	7	34	46	13	1			1.67	42	0.69
	2	57	B	3849	3062 (80)	11	36	37	14	1			1.57	39	0.72
Form P	1	6	B	3805	3369 (89)	13	14	47	24	2			1.88	47	0.71
	1	14	B	3805	3477 (91)	5	20	55	19	1			1.91	48	0.65
	1	20	B	3805	3617 (95)	1	12	62	24	1			2.12	53	0.52
	1	28	E	3805	2848 (75)	7	40	42	11	<1			1.58	40	0.69
	1	34	B	3805	2933 (77)	35	39	18	8	<1			0.99	25	0.66
	2	46	B	3805	2921 (77)	38	30	19	12	1			1.07	27	0.59
	2	52	B	3805	3281 (86)	17	37	34	12	1			1.44	36	0.73
	2	57	B	3805	2844 (75)	12	33	39	14	1			1.58	40	0.73
Form Q	1	6	B	3773	3542 (94)	2	15	55	26	1			2.09	52	0.62
	1	14	B	3773	3549 (94)	6	21	56	17	1			1.85	46	0.60
	1	20	B	3773	3397 (90)	5	29	51	15	<1			1.78	44	0.66
	1	28	E	3773	3187 (84)	2	30	51	16	<1			1.82	45	0.64
	1	34	B	3773	3236 (86)	8	28	48	16	1			1.72	43	0.66
	2	46	B	3773	3418 (91)	4	48	37	11	1			1.55	39	0.62
	2	52	B	3773	2945 (78)	6	27	50	17	1			1.80	45	0.68
	2	57	B	3773	3066 (81)	8	46	35	11	1			1.52	38	0.69
Form R	1	6	B	3726	3569 (96)	2	22	54	21	1			1.98	49	0.62
	1	14	B	3726	3425 (92)	3	25	56	15	1			1.86	47	0.62
	1	20	B	3726	3414 (92)	2	18	60	19	1			1.98	50	0.65

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.

Table A2. Summary Statistics for May Constructed Response Items (cont.)

	Session	Item Number	(B)CR / (E)CR	Number of Students in Item Analysis Sample	Number (and Percent) of Students with Scorable Responses	Percent of Students Receiving These Scores*						Mean Raw Score	Mean Percent	Item, Total Correlation**
						0	1	2	3	4	5			
Government														
Form R	1	28	E	3726	2732 (73)	7	28	46	19	<1		1.77	44	0.68
	1	34	B	3726	2990 (80)	19	35	35	11	<1		1.39	35	0.69
	2	46	B	3726	2824 (76)	19	32	32	16	<1		1.47	37	0.72
	2	52	B	3726	3281 (88)	4	32	48	16	<1		1.78	44	0.65
	2	57	B	3726	3129 (84)	6	30	46	17	<1		1.76	44	0.65
Form S	1	6	B	3644	3277 (90)	3	28	55	14	<1		1.81	45	0.55
	1	14	B	3644	3298 (91)	6	26	49	19	1		1.84	46	0.71
	1	20	B	3644	3382 (93)	2	21	55	21	<1		1.97	49	0.66
	1	28	E	3644	2777 (76)	6	31	40	22	<1		1.79	45	0.69
	1	34	B	3644	3025 (83)	13	31	38	16	1		1.61	40	0.74
	2	46	B	3644	3306 (91)	2	28	53	15	1		1.84	46	0.62
	2	52	B	3644	3261 (89)	2	22	59	18	<1		1.93	48	0.58
	2	57	B	3644	2967 (81)	3	28	53	15	1		1.81	45	0.64

*Students with condition codes are not included in these percentages. Also, the value "<1" indicates that the percent is greater than zero and less than 0.5.

**Item-total Correlation = correlation between the score on the item, and the total score, where the total score does not include the score on the studied item.