

Section 4. Test-Level Analyses

Test-level analyses include demographic distributions, reliability analyses, summary statistics, and decision consistency and accuracy.

Demographic Distributions

All eligible students completed the MDHSA. The demographic characteristics of the students are given in Tables 4.1 to 4.3 for the January and May administrations of Algebra, Biology, and Government, respectively. The number of students participating in the May administration was greater than the number of students participating in the January administration. As a result, only two field test versions were included in the January administration to ensure sufficient sample sizes for the analyses of the field test items. Due to the small numbers of students participating in the July administration, the May field test sections were repeated to ensure that the test length was comparable.

Table 4.1 Demographic Information for Algebra

		January Primary Forms		January Make-Up Forms		May Primary Forms		May Make-Up Forms	
		N	%	N	%	N	%	N	%
Overall		5921		904		68550		3221	
Gender									
	Male	2962	50.03	469	51.88	34150	49.82	1694	52.59
	Female	2870	48.47	428	47.35	34092	49.73	1414	43.90
	Missing	89	1.50	7	0.77	308	0.45	113	3.51
Special Education									
	Yes	31	0.52	4	0.44	643	0.94	42	1.30
	No	5888	99.44	900	99.56	67792	98.89	3174	98.54
	504	2	0.03	0	0.00	115	0.17	5	0.16
Ethnicity									
	American Indian	24	0.41	3	0.33	227	0.33	20	0.62
	Asian/Pacific Islander	153	2.58	21	2.32	3710	5.41	75	2.33
	African American	2487	42.00	584	64.60	24778	36.15	1456	45.20
	White	2786	47.05	221	24.45	34813	50.78	1267	39.34
	Hispanic	281	4.75	41	4.54	4226	6.16	179	5.56
	Missing	190	3.21	34	3.76	796	1.16	224	6.95
Limited English Proficient									
	Yes	0	0.00	0	0.00	138	0.20	6	0.19
	No	5921	100.00	904	100.00	68373	99.74	3215	99.81
	Exited	0	0.00	0	0.00	39	0.06	0	0.00

Table 4.2 Demographic Information for Biology

		January Primary Forms		January Make-Up Forms		May Primary Forms		May Make-Up Forms	
		N	%	N	%	N	%	N	%
Overall		7533		866		68550		2592	
Gender									
	Male	3660	48.59	435	50.23	34150	49.82	1331	51.35
	Female	3824	50.76	409	47.23	34092	49.73	1228	47.38
	Missing	49	0.65	22	2.54	308	0.45	33	1.27
Special Education									
	Yes	36	0.48	3	0.35	643	0.94	34	1.31
	No	7497	99.52	863	99.65	67792	98.89	2553	98.50
	504	0	0.00	0	0.00	115	0.17	5	0.19
Ethnicity									
	American Indian	18	0.24	4	0.46	227	0.33	13	0.50
	Asian/Pacific Islander	156	2.07	24	2.77	3710	5.41	86	3.32
	African American	1890	25.09	470	54.27	24778	36.15	1181	45.56
	White	5129	68.09	281	32.45	34813	50.78	1034	39.89
	Hispanic	255	3.39	56	6.47	4226	6.16	184	7.10
	Missing	85	1.13	31	3.58	796	1.16	94	3.63
Limited English Proficient									
	Yes	0	0.00	0	0.00	138	0.20	4	0.15
	No	7533	100.00	866	100.00	68373	99.74	2587	99.81
	Exited	0	0.00	0	0.00	39	0.06	1	0.04

Table 4.3 Demographic Information for Government

		January Primary Forms		January Make-Up Forms		May Primary Forms		May Make-Up Forms	
		N	%	N	%	N	%	N	%
Overall		7648		849		54349		2252	
Gender									
	Male	3745	48.97	410	48.29	26763	49.24	1172	52.04
	Female	3867	50.56	408	48.06	27375	50.37	1035	45.96
	Missing	36	0.47	31	3.65	211	0.39	45	2.00
Special Education									
	Yes	14	0.18	5	0.59	576	1.06	23	1.02
	No	7634	99.82	844	99.41	53676	98.76	2227	98.89
	504	0	0.00	0	0.00	97	0.18	2	0.09
Ethnicity									
	American Indian	21	0.27	5	0.59	205	0.38	12	0.53
	Asian/Pacific Islander	132	1.73	18	2.12	3346	6.16	64	2.84
	African American	2153	28.15	486	57.24	19326	35.56	921	40.90
	White	5033	65.81	247	29.09	27499	50.60	971	43.12
	Hispanic	230	3.01	45	5.30	3558	6.55	174	7.73
	Missing	79	1.03	48	5.65	415	0.76	110	4.88
Limited English Proficient									
	Yes	0	0.00	0	0.00	100	0.18	4	0.18
	No	7648	100.00	849	100.00	54218	99.76	2247	99.78
	Exited	0	0.00	0	0.00	0	0.00	0	0.00

Reliability

Reliability describes the extent to which differences in test scores reflect true differences in the knowledge, ability, or skill being tested rather than fluctuations due to chance or factors other than those which were being tested. The variance in the distributions of test scores (i.e., the differences among individuals) is partly due to real differences in the knowledge, skill, or ability being tested (true-score variance) and partly due to random errors in the measurement process (error variance).

The number used to describe reliability is an estimate of the proportion of the total variance that is true variance. Several different ways of estimating this proportion exist. The estimates of reliability given in this report are internal-consistency measures, which were derived from analysis of the consistency of the performance of individuals on items within a test (internal-consistency reliability). Therefore, the estimates apply only to the test form being analyzed. They do not take into account form-to-form variation due to equating limitations or lack of parallelism, nor were they responsive to day-to-day variation due to, for example, state of health or testing environment. Reliability coefficients range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely individuals would be to obtain very similar scores if they took another form of the test.

The formula for the internal consistency reliability as measured by Cronbach's alpha (Cronbach, 1951) is reported below:

$$\alpha = \frac{n}{n-1} \left[1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma_x^2} \right],$$

where n is the number of items, σ_i^2 is the variance of scores on the i -th item, and σ_x^2 is the variance of the total score (sum of scores on the individual items).

Since all four content areas contain mixed item types (dichotomous and polytomous items), it is more appropriate to report stratified alpha (Feldt & Brennan, 1989). Stratified alpha is a weighted average of Cronbach's alpha for item sets with different maximum score points or "strata." The formula for calculating the stratified alpha is:

$$strata \rho = 1 - \frac{\sum \sigma_{x_j}^2 (1 - \alpha_j)}{\sigma_x^2}$$

where $\sigma_{x_j}^2$ is the variance for strata j of the test, σ_x^2 is the total variance of the test, and α_j is the Cronbach's alpha for strata j of the test.

The results for the reliability analyses of the total test score are presented with the summary statistics in Tables 4.9 to 4.16. The reliability results indicate that all of the MDHSAs were highly reliable: reliabilities ranged from 0.89 to 0.95 for the primary forms, and from 0.90 to 0.96 for the make-up forms.

Summary Statistics

The mean scale scores by content area for the January and May administrations are presented in Table 4.4. The mean scores for Algebra, Biology, and Government were higher for the May administration compared to the January administration.

Table 4.4 Mean Scores by Administration

	Jan 2006			May 2006		
	N	Mean	SD	N	Mean	SD
Algebra	6825	404.18	46.39	71771	423.38	41.56
Biology	8399	407.55	43.73	52914	416.31	40.01
Government	8497	407.36	41.35	56601	420.40	46.29

The 2006 data presented in Tables 4.5 to 4.7 are based on students who took the tests in January, May, and July. The mean scale scores are presented for the years 2003 to 2006 by content area in Table 4.5. The mean scale scores are highest for 2006 compared to the other years.

Table 4.5 Comparison of Mean Scores

	2003	2004	2005	2006
Algebra	408.3	411.9	409.5	421.5
Biology	400.8	406.2	404.7	415.0
Government	403.5	406.5	409.3	418.5

Table 4.6 presents the passing rates for Algebra, Biology, and Government. The passing rate for all three content areas is highest for 2006 compared to the passing rates for 2003, 2004, and 2005. As can be seen in the table, 66.6% of students achieved the proficient standard or above in the three subject areas.

Table 4.6 Comparison of Percentage Passing Rates

	2003	2004	2005	2006
Algebra	53.1	59.3	54.5	66.6
Biology	54.3	62.0	58.4	67.7
Government	39.8	54.6	67.1	74.1

Table 4.7 presents the percent of Algebra students classified as basic, proficient, and advanced. Algebra is the mathematics component of the MSDE Adequate Yearly Progress report as required under the NCLB act. As can be seen in the table, one-third of

students achieved the Basic performance level, whereas two-thirds of students achieved the Proficient or Advanced performance levels.

Table 4.7 Classification Rates for Algebra

	2006
Basic	33.4
Proficient	40.7
Advanced	25.9

Summary statistics for all students and for subgroups based on gender, special education programs, ethnicity, and English language fluency are presented in Tables 4.8 through 4.13. The tables include number of students tested for whom valid scores were available, mean scale scores, and standard deviations of scale scores. In addition, test score reliabilities are provided for the overall group of examinees. Information is presented for the primary forms of the content area, followed by the make-up forms.

Results indicate that for each administration and content area, students who took the primary form scored higher on average than students who took the make-up form(s). In addition, performance varied over subgroups. At the overall examinee level, score reliability coefficients ranged from 0.89 (May primary algebra form) to 0.95 (May primary government form).

Table 4.8 Summary Statistics for Algebra Primary Forms

		January				May			
		Mean	SD	N	Alpha	Mean	SD	N	Alpha
Overall		407.86	45.42	5921	0.90	424.94	40.36	68550	0.89
Gender									
	Male	403.54	49.79	2962		422.98	43.44	34150	
	Female	412.42	39.95	2870		427.17	36.65	34092	
	Missing	404.21	44.60	89		395.18	52.89	308	
Special Education									
	Yes	*	*	31		394.12	42.73	643	
	No	408.02	45.39	5888		425.21	40.24	67792	
	504	*	*	2		433.03	27.36	115	
Ethnicity									
	American Indian	*	*	24		417.34	43.48	227	
	Asian/Pacific Islander	421.61	35.57	153		447.27	31.98	3710	
	African American	385.38	47.58	2487		405.88	41.59	24778	
	White	428.89	32.86	2786		437.83	33.00	34813	
	Hispanic	396.11	38.51	281		415.63	40.96	4226	
	Missing	399.04	44.01	190		401.70	51.88	796	
Limited English Proficient									
	Yes	*	*	0		380.36	58.41	138	
	No	407.86	45.42	5921		425.03	40.27	68373	
	Exited	*	*	0		*	*	39	

* Statistics not reported for sample size less than 50

Table 4.9 Summary Statistics for Algebra Make Up Forms

		January Make-Up Forms								May Make-Up Forms							
		C				D				X				Y			
		Mean	SD	N	Alpha	Mean	SD	N	Alpha	Mean	SD	N	Alpha	Mean	SD	N	Alpha
Overall		384.38	44.68	588	0.91	372.19	46.04	316	0.92	391.74	51.39	2680	0.92	382.90	52.17	541	0.91
Gender																	
	Male	381.24	47.87	303		367.43	46.23	166		386.54	56.52	1400		378.17	54.00	294	
	Female	387.5	40.96	282		378.14	45.75	146		398.62	43.93	1182		389.25	49.56	232	
	Missing	*	*	3		*	*	4		382.83	47.98	98		*	*	15	
Special Education																	
	Yes	*	*	4		*	*	0		*	*	39		*	*	3	
	No	384.41	44.79	584		372.19	46.04	316		391.94	51.45	2637		383.03	52.26	537	
	504	*	*	0		*	*	0		*	*	4		*	*	1	
Ethnicity																	
	American Indian	*	*	2		*	*	1		*	*	15		*	*	5	
	Asian/Pacific Islander	*	*	12		*	*	9		425.32	45.14	65		*	*	10	
	African American	372.73	42.71	386		363.87	43.35	198		377.31	49.92	1183		372.91	50.88	273	
	White	410.57	38.48	155		397.68	43.86	66		407.86	48.89	1075		400.10	51.92	192	
	Hispanic	*	*	19		*	*	22		383.43	51.01	152		*	*	27	
	Missing	*	*	14		*	*	20		386.45	45.90	190		*	*	34	
Limited English Proficient																	
	Yes	*	*	0		*	*	0		*	*	6		*	*	0	
	No	384.38	44.68	588		372.19	46.04	316		391.80	51.25	2674		382.90	52.17	541	
	Exited	*	*	0		*	*	0		*	*	0		*	*	0	

* Statistics not reported for sample size less than 50 (N<50)

Table 4.10 Summary Statistics for Biology Primary Forms

		January				May			
		Mean	SD	N	Alpha	Mean	SD	N	Alpha
Overall		411.16	41.96	7533	0.94	417.93	39.04	50322	0.94
Gender									
	Male	407.17	45.27	3660		415.77	42.33	24727	
	Female	415.48	37.83	3824		420.14	35.26	25380	
	Missing	*	*	49		404.01	50.30	215	
Special Education									
	Yes	*	*	36		383.52	40.18	483	
	No	411.33	41.95	7497		418.27	38.88	49772	
	504	*	*	0		413.33	33.10	67	
Ethnicity									
	American Indian	*	*	18		408.41	40.00	188	
	Asian/Pacific Islander	425.06	40.55	156		438.99	37.13	3033	
	African American	381.38	42.77	1890		399.38	36.19	16974	
	White	422.93	35.46	5129		429.01	35.81	26617	
	Hispanic	399.41	40.14	255		406.21	36.03	3117	
	Missing	372.61	34.41	85		403.11	45.59	393	
Limited English Proficient									
	Yes	*	*	0		386.19	39.64	57	
	No	411.16	41.96	7533		417.96	39.03	50247	
	Exited	*	*	0		*	*	18	

* Statistics not reported for sample size less than 50

Table 4.11 Summary Statistics for Biology Make Up Forms

		January Make-Up Forms								May Make-Up Forms							
		C				D				X				Y			
		Mean	SD	N	Alpha	Mean	SD	N	Alpha	Mean	SD	N	Alpha	Mean	SD	N	Alpha
Overall		381.39	45.38	571	0.93	366.06	46.17	295	0.90	385.42	46.11	2271	0.92	381.66	40.30	321	0.90
Gender																	
	Male	379.80	50.81	292		358.83	49.79	143		378.95	51.21	1177		378.37	41.79	154	
	Female	383.97	39.06	267		374.08	41.02	142		392.78	38.04	1069		385.70	37.40	159	
	Missing	*	*	12		*	*	10		*	*	25		*	*	8	
Special Education																	
	Yes	*	*	3		*	*	0		*	*	23		*	*	11	
	No	381.28	45.46	568		366.06	46.17	295		385.75	45.91	2244		382.69	39.54	309	
	504	*	*	0		*	*	0		*	*	4		*	*	1	
Ethnicity																	
	American Indian	*	*	3		*	*	1		*	*	13		*	*	0	
	Asian/Pacific Islander	*	*	9		*	*	15		410.99	43.26	78		*	*	8	
	African American	363.64	40.34	311		354.06	42.36	159		374.03	41.95	1031		378.85	33.90	150	
	White	404.63	39.32	208		390.15	46.21	73		395.52	45.47	926		385.69	47.94	108	
	Hispanic	*	*	24		*	*	32		380.25	45.86	153		*	*	31	
	Missing	*	*	16		*	*	15		408.94	58.38	70		*	*	24	
Limited English Proficient																	
	Yes	*	*	0		*	*	0		*	*	3		*	*	1	
	No	381.39	45.38	571		366.06	46.17	295		385.47	46.14	2267		381.47	40.22	320	
	Exited	*	*	0		*	*	0		*	*	1		*	*	0	

* Statistics not reported for sample size less than 50 (N<50)

Table 4.12 Summary Statistics for Government Primary Forms

		January				May			
		Mean	SD	N	Alpha	Mean	SD	N	Alpha
Overall		410.61	39.92	7648	0.94	421.53	45.55	54349	0.95
Gender									
	Male	408.86	42.06	3745		419.01	48.39	26763	
	Female	412.63	37.54	3867		424.15	42.32	27375	
	Missing	*	*	36		401.98	54.50	211	
Special Education									
	Yes	*	*	14		392.53	36.35	576	
	No	410.68	39.90	7634		421.85	45.55	53676	
	504	*	*	0		414.71	37.44	97	
Ethnicity									
	American Indian	*	*	21		414.38	36.26	205	
	Asian/Pacific Islander	421.40	36.41	132		442.30	46.23	3346	
	African American	383.55	39.00	2153		401.89	41.12	19326	
	White	423.11	33.62	5033		434.89	42.96	27499	
	Hispanic	399.37	36.88	230		408.51	41.58	3558	
	Missing	367.04	45.86	79		399.14	52.36	415	
Limited English Proficient									
	Yes	*	*	0		390.68	49.82	100	
	No	410.61	39.92	7648		421.59	45.53	54218	
	Exited	*	*	0		*	*	31	

* Statistics not reported for sample size less than 50

Table 4.13 Summary Statistics for Government Make Up Forms

		January Make-Up Forms								May Make-Up Forms							
		C				D				X				Y			
		Mean	SD	N	Alpha	Mean	SD	N	Alpha	Mean	SD	N	Alpha	Mean	SD	N	Alpha
Overall		381.83	41.35	576	0.93	370.05	43.89	273	0.92	395.44	55.00	1944	0.96	377.86	51.22	308	0.94
Gender																	
	Male	374.92	47.40	262		367.98	49.33	148		389.01	58.20	1012		375.49	52.75	160	
	Female	388.74	34.98	293		373.97	35.05	115		403.71	50.20	891		380.86	48.39	144	
	Missing	*	*	21		*	*	10		*	*	41		*	*	4	
Special Education																	
	Yes	*	*	5		*	*	0		*	*	17		*	*	6	
	No	381.98	41.42	571		370.05	43.89	273		395.84	54.81	1927		378.56	50.53	300	
	504	*	*	0		*	*	0		*	*	0		*	*	2	
Ethnicity																	
	American Indian	*	*	2		*	*	3		*	*	10		*	*	2	
	Asian/Pacific Islander	*	*	12		*	*	6		418.51	43.15	53		*	*	11	
	African American	373.37	39.57	322		362.24	42.43	164		375.30	50.54	761		370.89	49.64	160	
	White	400.30	34.38	184		391.63	43.88	63		415.97	54.04	878		390.32	52.53	93	
	Hispanic	*	*	25		*	*	20		380.99	45.15	149		*	*	25	
	Missing	*	*	31		*	*	17		378.86	43.91	93		*	*	17	
Limited English Proficient																	
	Yes	*	*	0		*	*	0		*	*	4		*	*	0	
	No	381.83	41.35	576		370.05	43.89	273		395.49	55.05	1939		377.86	51.22	308	
	Exited	*	*	0		*	*	0		*	*	1		*	*	0	

* Statistics not reported for sample size less than 50 (N<50)

Decision Accuracy and Consistency

The accuracy of decisions based on specified cut-scores was assessed for Reliability of Classification using the computer program RELCLASS, which is an ETS proprietary software. RELCLASS provides two statistics that describe the reliability of classifications based on test scores (Livingston & Lewis, 1995). More specifically, information from an administration of one form is used to estimate the following:

- 1) Decision Accuracy describes the extent to which examinees are classified in the same way as they would be on the basis of the average of all possible forms of a test. Decision accuracy answers the question: How does the actual classification of test takers, based on their single-form scores, agree with the classification that would be made on the basis of their true scores, if their true scores were somehow known?
- 2) Decision Consistency describes the extent to which examinees are classified in the same way as they would be on the basis of a single form of a test other than the one for which data are available. Decision consistency answers the question: What is the agreement between the classifications based on two non-overlapping, equally difficult forms of the test?

RELCLASS estimates decision accuracy using an estimated joint distribution of reported performance level classifications on the current form of the exam and the performance level classifications based on an all-forms average (true score). RELCLASS estimates decision consistency using an estimated joint distribution of reported performance level classifications on the current form of the exam and performance level classifications on the alternate (parallel) form. In each case, the proportion of performance level classifications with exact agreement is the sum of the entries in the diagonal of the contingency table representing the joint distribution.

The results are provided in Tables 4.14 to 4.16 by content area. The tables show that decision accuracy values ranged from 0.82 to 0.94, across all performance levels and from 0.89 to 0.90 for the Proficient and Above classification in algebra. Therefore, the agreement between classifications based on an observable variable (scores on one form of a test) and classifications based on an unobservable variable (the test takers' true scores) was very good. Decision consistency values ranged from 0.73 to 0.92 across all performance levels, and from 0.85 to 0.86 for the Proficient and Above classifications in algebra. Since decision consistency statistics describe the agreement between classifications based on two variables (scores on the form students have taken and a parallel form of the same test that is not administered to the students), these values are within the acceptable range.

Note that in all cases the decision accuracy indices are somewhat larger than the decision consistency indices. This is due to differences in the estimation procedures. The estimation procedure for decision accuracy includes a random component on one of the two variables, whereas in estimating decision consistency each variable includes a random component (Livingston & Lewis, 1995).

Table 4.14 Decision Accuracy and Consistency for the Algebra Tests

	Placement Score	Advanced	Proficient	Basic	Category Total*
January, 2006					
Decision Accuracy	450 – 515	0.12	0.02	0.00	0.14
	412 – 449	0.05	0.24	0.08	0.36
	240 – 411	0.00	0.03	0.47	0.50
	Estimated Proportion Correctly Classified*: Total = 0.82; Proficient & Above = 0.89				
Decision Consistency	450 – 515	0.11	0.03	0.00	0.14
	412 – 449	0.07	0.20	0.09	0.36
	240 – 411	0.00	0.05	0.45	0.50
	Estimated Proportion Correctly Classified*: Total =0.76; Proficient & Above = 0.86				
May, 2006					
Decision Accuracy	450 – 650	0.21	0.06	0.00	0.27
	412 – 449	0.05	0.29	0.08	0.41
	240 – 411	0.00	0.01	0.30	0.32
	Estimated Proportion Correctly Classified*: Total = 0.80; Proficient & Above = 0.90				
Decision Consistency	450 – 650	0.20	0.07	0.01	0.27
	412 – 449	0.07	0.24	0.10	0.41
	240 – 411	0.00	0.03	0.29	0.32
	Estimated Proportion Correctly Classified*: Total =0.73; Proficient & Above = 0.87				

*Inconsistencies with category cell entries are due to rounding.

Table 4.15 Decision Accuracy and Consistency for the Biology Tests

	Placement Score	Proficient	Basic	Category Total*
January, 2006				
Decision Accuracy	400 – 546	0.56	0.06	0.62
	240 – 399	0.02	0.36	0.38
	Estimated Proportion Correctly Classified*: Total = 0.92			
Decision Consistency	400 – 546	0.54	0.08	0.62
	240 – 399	0.03	0.35	0.38
	Estimated Proportion Correctly Classified*: Total =0.89			
May, 2006				
Decision Accuracy	400 – 588	0.67	0.02	0.69
	240 – 399	0.04	0.27	0.31
	Estimated Proportion Correctly Classified*: Total = 0.94			
Decision Consistency	400 – 588	0.66	0.03	0.69
	240 – 399	0.05	0.27	0.31
	Estimated Proportion Correctly Classified*: Total =0.92			

*Inconsistencies with category cell entries are due to rounding.

Table 4.16 Decision Accuracy and Consistency for the Government Tests

	Placement Score	Proficient	Basic	Category Total*
January, 2006				
Decision Accuracy	394 – 606	0.65	0.03	0.68
	240 – 393	0.03	0.29	0.32
	Estimated Proportion Correctly Classified*: Total = 0.94			
Decision Consistency	394 – 606	0.64	0.04	0.68
	240 – 393	0.04	0.28	0.32
	Estimated Proportion Correctly Classified*: Total =0.92			
May, 2006				
Decision Accuracy	394 – 650	0.71	0.04	0.75
	240 – 393	0.01	0.24	0.25
	Estimated Proportion Correctly Classified*: Total = 0.95			
Decision Consistency	394 – 650	0.70	0.06	0.75
	240 – 393	0.02	0.23	0.25
	Estimated Proportion Correctly Classified*: Total =0.92			

*Inconsistencies with category cell entries are due to rounding.