Section 10. Validity

Validity is one of the most important attributes of assessment quality. Validity refers to the degree to which logical, empirical, and judgmental evidence supports each proposed interpretation or use of a set of scores, and it is one of the most fundamental considerations in developing and evaluating tests (AERA, APA, & NCME, 1999; Messick, 1989). Validity is not based on a single study or type of study but is an ongoing process of gathering evidence supporting the interpretation or use of the resulting test scores. The process begins with the test design and continues throughout the entire assessment process, including design, content specifications, item development, psychometric quality, and inferences made from the test results.

Students' scores on an MD Mod-HSA are inferred to reflect students' level of knowledge and skills in a content area. The scores are used to classify students in terms of their level of proficiency, based on cut-scores established by the state.

Evidence Based on Analyses of Test Content

The Maryland Mod-HSAs are referred to as end-of-course tests because students take each test as they complete the appropriate coursework. Banked HSA items were selected and adapted for the MD Mod-HSAs to measure the knowledge and skills expected of students following completion of coursework.

The constructs measured by each MD Mod-HSA are described in detail in the Maryland high school curriculum standards, or Core Learning Goals. All ETS content staff working on item selection and development have been trained in the CLGs. The test blueprint documents presented in Section 9 (see Tables 9.1 to 9.4) were created in collaboration with committees of Maryland educators and were derived from the Maryland goals, expectations, and indicators.

The process of selecting and adapting banked MD HSA items for use as MD Mod-HSA items is summarized briefly in Section 9 and described in detail in the Maryland Modified High School Assessment 2008 Technical Report.[12] Banked items were referenced to a particular instructional standard (i.e., goal, expectation, or indicator). During the internal ETS development process, the specific reference was confirmed or changed to reflect changes to the item. When the item went to a committee of Maryland educators for content review, the members of the committee made independent judgments about the match of the item content to the standard it was intended to measure and evaluated the appropriateness for the age and cognitive ability of the students being tested.

[12] Available at http://marylandpublicschools.org/MSDE/divisions/planningresultstest/HSA+Technical+Reports.htm.

Evidence Based on Analyses of Internal Test Structure

Exploratory Factor Analysis

To investigate the dimensionality of the MD Mod-HSA operational forms, exploratory factor analyses were conducted at the item level for each 50-item operational form created after the May 2008 test administration. The software program *Mplus* (Muthén & Muthén, 2007) was used to generate tetrachoric correlations that were then read into the program for the analyses. The estimator used in these exploratory analyses was a weighted least-squares with mean and variance adjustment (Muthén, DuToit, & Spisic, 1997). This estimator was specifically designed for the analysis of ordered categorical data. Solutions were rotated by Quartimin methods because the factors were expected to be correlated.

Two groups of students took the MD Mod-HSAs in May 2008; data from each group were analyzed separately for the exploratory factor analyses. The first group was the *target* population, made up of students identified by MSDE as being eligible to take the MD Mod-HSAs. These students took the MD Mod-HSA instead of the regular MD HSA. The second group of students was the *linking* sample, which consisted of regular MD HSA examinees identified by MSDE to take the MD Mod-HSA in the same content area as their May MD HSA. The data provided by this second group of examinees were used to calibrate the MD Mod-HSA forms and to align these forms to the MD HSA reporting scale.

The percentage of score variance accounted for by each factor having an eigenvalue greater than 1.0 is shown in Tables 10.1 to 10.8 for each form. The decision to include only eigenvalues greater than 1.0 follows the Kaiser-Guttman rule (Kaiser, 1960). Scree plots (Catell, 1966) for each form are given in Figures 10.1 to 10.16 for the first 50 factors extracted. The scree plot involves plotting the eigenvalues of the factors extracted in order of magnitude from high to low. The plot is examined for a point at which the decrease in eigenvalues levels off. Factors prior to this point are considered important because of the variance they explain. Factors at and beyond this point add relatively little information.

Examination of the plots and tables for the linking sample shows that the eigenvalues for the first factors ranged from about 12.0 to 15.6 across forms and subject areas, and these first factors accounted for 24 to 31 percent of the variance. The eigenvalues for the second and subsequent factors were no greater than about 2.0, and these factors accounted for about 2 to 5 percent of the remaining variance. Results for the two forms taken by the linking sample were very similar across forms. The sizable amount of variance accounted for by the first factor indicates a large first factor; confirmatory factor analyses or a study of the essential dimensionality of the data for the linking sample could be used to assess the fit of a single factor model to the data.

With regard to the target population, the first factor results tended to be about half of those obtained for the linking sample. Specifically, the eigenvalues for the first factors ranged from about 5.5 to 7.0, and this factor accounted for about 11 to 14 percent of the variance. Thus, for the target population a much smaller first factor was found. Like the linking sample, the second and subsequent factors had small eigenvalues and accounted for 4 percent or less of the remaining variance.

The lower eigenvalues and percentages of score variance accounted for by the first factor in the target population appears to be a product of the difficulty of the MD Mod-HSA items for students in the target population. Table 10.9 shows that for the linking sample the mean item p-values were in the low 0.70s, a moderate degree of difficulty. For the target population, Table 10.10 shows that p-values were in the high 0.40s, on average. The MD Mod-HSA items are multiple-choice items with three answer choices; therefore the item p-values could reflect a considerable amount of guessing.

Very difficult items discriminate less well than do moderately difficult items and introduce more error because of increased guessing. As shown in Tables 10.11 and 10.12, the MD Mod-HSA item point-biserials were considerably lower for the target population than they were for the linking sample. Also the internal consistency results were notably lower: for the linking sample, internal consistency ranged from 0.86 to 0.89 across subject areas, whereas for the target population it ranged from 0.71 to 0.79. Comparison of the tetrachoric correlations read into the factor analyses and summarized in Table 10.13 also shows that the item intercorrelations for the target population were quite low and about half the size of those observed for the linking sample, on average. Presumably as achievement in the target population improves, item discrimination, internal consistency, and the item intercorrelations will improve concomitantly.

Table 10.1 Factor Analysis Results for MD Mod-HSA Algebra, May 2008 Linking Sample

	Form 108		Form 208	
Factor	Eigenvalue	%Var	Eigenvalue	%Var
1	11.87	23.75	12.49	24.98
2	2.13	4.26	2.01	4.01
3	1.64	3.28	1.67	3.34
4	1.47	2.93	1.62	3.24
5	1.40	2.79	1.46	2.93
6	1.36	2.72	1.40	2.81
7	1.33	2.67	1.34	2.68
8	1.26	2.51	1.26	2.52
9	1.23	2.46	1.23	2.46
10	1.16	2.32	1.16	2.32
11	1.10	2.20	1.09	2.19
12	1.08	2.16	1.07	2.14
13	1.07	2.13	1.06	2.11
14	1.04	2.08	1.05	2.10
15	1.01	2.01		

Table 10.2 Factor Analysis Results for MD Mod-HSA Algebra, May 2008 Target Population

Factor	Form 108		Form 208	
	Eigenvalue	%Var	Eigenvalue	%Var
1	7.08	14.17	6.48	12.96
2	1.71	3.41	2.30	4.60
3	1.66	3.31	1.81	3.62
4	1.49	2.98	1.54	3.08
5	1.43	2.86	1.47	2.95
6	1.38	2.77	1.45	2.90
7	1.34	2.68	1.40	2.80
8	1.32	2.65	1.35	2.70
9	1.26	2.52	1.31	2.63
10	1.25	2.49	1.29	2.58
11	1.21	2.42	1.26	2.52
12	1.17	2.34	1.22	2.44
13	1.11	2.22	1.18	2.36
14	1.10	2.20	1.12	2.24
15	1.07	2.14	1.11	2.22
16	1.04	2.08	1.09	2.18
17	1.01	2.02	1.08	2.16
18			1.07	2.14
19			1.01	2.03

Table 10.3 Factor Analysis Results for MD Mod-HSA Biology, May 2008 Linking Sample

Factor	Form 108		Form 208	
	Eigenvalue	%Var	Eigenvalue	%Var
1	12.28	24.57	12.04	24.07
2	1.54	3.08	1.77	3.53
3	1.50	2.99	1.56	3.12
4	1.42	2.84	1.52	3.05
5	1.39	2.78	1.32	2.63
6	1.35	2.71	1.29	2.58
7	1.29	2.58	1.21	2.42
8	1.26	2.51	1.19	2.38
9	1.18	2.36	1.17	2.34
10	1.13	2.27	1.12	2.24
11	1.10	2.20	1.11	2.23
12	1.08	2.16	1.08	2.16
13	1.07	2.14	1.06	2.13
14	1.04	2.09	1.03	2.05
15	1.00	2.00	1.01	2.02

Table 10.4 Factor Analysis Results for MD Mod-HSA Biology, May 2008 Target Population

Factor	Form 108		Form 208	
	Eigenvalue	%Var	Eigenvalue	%Var
1	5.55	1.11	6.23	12.45
2	1.81	3.61	1.79	3.58
3	1.63	3.27	1.69	3.38
4	1.59	3.17	1.53	3.05
5	1.50	2.99	1.48	2.96
6	1.49	2.97	1.45	2.90
7	1.43	2.87	1.42	2.84
8	1.40	2.79	1.36	2.72
9	1.36	2.73	1.33	2.66
10	1.33	2.65	1.29	2.59
11	1.27	2.54	1.26	2.52
12	1.24	2.49	1.24	2.47
13	1.22	2.43	1.22	2.44
14	1.19	2.39	1.20	2.40
15	1.17	2.34	1.14	2.28
16	1.14	2.29	1.13	2.26
17	1.13	2.26	1.11	2.22
18	1.11	2.22	1.10	2.19
19	1.08	2.15	1.05	2.10
20	1.04	2.08		
21	1.01	2.02		

Table 10.5 Factor Analysis Results for MD Mod-HSA English, May 2008 Linking Sample

Factor	Form 108		Form 208	
	Eigenvalue	%Var	Eigenvalue	%Var
1	12.90	25.79	12.62	25.23
2	1.90	3.81	2.01	4.01
3	1.83	3.65	1.76	3.52
4	1.57	3.14	1.68	3.36
5	1.52	3.04	1.57	3.14
6	1.47	2.95	1.43	2.87
7	1.39	2.77	1.32	2.64
8	1.25	2.51	1.27	2.54
9	1.22	2.44	1.24	2.48
10	1.19	2.38	1.22	2.44
11	1.18	2.36	1.18	2.36
12	1.17	2.33	1.14	2.28
13	1.06	2.12	1.10	2.19
14	1.01	2.03	1.07	2.14
15			1.04	2.07

Table 10.6 Factor Analysis Results for MD Mod-HSA English, May 2008 Target Population

Factor	Form 108		Form 208	
	Eigenvalue	%Var	Eigenvalue	%Var
1	5.67	11.35	6.69	13.37
2	1.81	3.61	1.84	3.68
3	1.66	3.33	1.78	3.56
4	1.56	3.11	1.65	3.31
5	1.49	2.98	1.44	2.88
6	1.47	2.94	1.42	2.84
7	1.38	2.75	1.39	2.78
8	1.31	2.62	1.30	2.61
9	1.27	2.53	1.28	2.57
10	1.25	2.49	1.26	2.52
11	1.22	2.44	1.24	2.49
12	1.21	2.41	1.20	2.40
13	1.18	2.36	1.18	2.35
14	1.17	2.34	1.15	2.30
15	1.15	2.30	1.09	2.19
16	1.09	2.18	1.08	2.16
17	1.06	2.11	1.04	2.08
18	1.03	2.06	1.01	2.03
19	1.01	2.02	1.01	2.01

Table 10.7 Factor Analysis Results for MD Mod-HSA Government, May 2008 Linking Sample

Factor	Form 108		Form 208	
	Eigenvalue	%Var	Eigenvalue	%Var
1	14.74	29.47	15.61	31.22
2	1.93	3.86	1.74	3.48
3	1.66	3.31	1.60	3.20
4	1.45	2.89	1.51	3.02
5	1.41	2.81	1.40	2.79
6	1.34	2.69	1.30	2.60
7	1.29	2.58	1.25	2.50
8	1.25	2.49	1.20	2.39
9	1.18	2.36	1.15	2.31
10	1.11	2.22	1.12	2.23
11	1.08	2.16	1.09	2.17
12	1.06	2.12	1.05	2.10
13	1.01	2.02	1.01	2.02

Table 10.8 Factor Analysis Results for MD Mod-HSA Government, May 2008 Target Population

	Form 108		Form 208	
Factor	Eigenvalue	%Var	Eigenvalue	%Var
1	6.50	12.99	6.58	13.16
2	1.86	3.71	1.90	3.81
3	1.55	3.10	1.77	3.55
4	1.51	3.02	1.66	3.31
5	1.42	2.84	1.52	3.04
6	1.38	2.77	1.41	2.83
7	1.37	2.73	1.39	2.78
8	1.33	2.67	1.33	2.67
9	1.30	2.60	1.32	2.63
10	1.25	2.50	1.27	2.54
11	1.22	2.44	1.26	2.52
12	1.20	2.39	1.21	2.42
13	1.16	2.31	1.16	2.32
14	1.14	2.29	1.14	2.28
15	1.09	2.19	1.10	2.21
16	1.08	2.17	1.07	2.15
17	1.06	2.12	1.04	2.08
18	1.04	2.07	1.02	2.05
19	1.02	2.05	1.02	2.05

Table 10.9 Distributions of P-Values: MD Mod-HSA May 2008 Operational Items—Linking Sample

	Number and Percentage of Items							
	Algebra		Biology		English		Government	
P-Value	N	%	N	%	N	%	N	%
$P < 0.10$	0	0.0	0	0.0	0	0.0	0	0.0
$0.10 \leq P < 0.20$	0	0.0	0	0.0	0	0.0	0	0.0
$0.20 \leq P < 0.30$	1	1.0	1	1.0	0	0.0	0	0.0
$0.30 \leq P < 0.40$	1	1.0	1	1.0	1	1.0	0	0.0
$0.40 \leq P < 0.50$	4	4.0	6	6.1	3	3.1	2	2.0
$0.50 \leq P < 0.60$	12	12.0	13	13.3	9	9.3	10	10.0
$0.60 \leq P < 0.70$	22	22.0	20	20.4	16	16.5	17	17.0
$0.70 \leq P < 0.80$	28	28.0	27	27.6	26	26.8	33	33.0
$0.80 \leq P < 0.90$	26	26.0	24	24.5	34	35.1	29	29.0
$P \geq 0.90$	6	6.0	6	6.1	8	8.3	9	9.0
Descriptive Statistics								
N Items*	100		98		97		100	
Mean	0.72		0.71		0.74		0.75	
SD	0.14		0.14		0.13		0.11	
Min	0.29		0.23		0.31		0.48	
Max	0.96		0.93		0.93		0.94	

* Includes the number of unique items; some Biology and English items appear on both Form 108 and Form 208.

Table 10.10 Distributions of P-Values: MD Mod-HSA May 2008 Operational Items—Target Population

	Number and Percentage of Items							
	Algebra		Biology		English		Government	
P-Value	N	%	N	%	N	%	N	%
$P < 0.10$	0	0.0	0	0.0	0	0.0	0	0.0
$0.10 \leq P < 0.20$	0	0.0	1	1.0	1	1.0	0	0.0
$0.20 \leq P < 0.30$	7	7.0	7	7.1	1	1.0	4	4.0
$0.30 \leq P < 0.40$	21	21.0	20	20.4	19	19.6	21	21.0
$0.40 \leq P < 0.50$	32	32.0	27	27.6	27	27.8	36	36.0
$0.50 \leq P < 0.60$	22	22.0	22	22.5	24	24.7	23	23.0
$0.60 \leq P < 0.70$	13	13.0	17	17.4	21	21.7	13	13.0
$0.70 \leq P < 0.80$	5	5.0	4	4.1	4	4.1	3	3.0
$0.80 \leq P < 0.90$	0	0.0	0	0.0	0	0.0	0	0.0
$P \geq 0.90$	0	0.0	0	0.0	0	0.0	0	0.0
Descriptive Statistics								
N Items*	100		98		97		100	
Mean	0.47		0.48		0.50		0.48	
SD	0.13		0.13		0.12		0.11	
Min	0.23		0.16		0.18		0.23	
Max	0.79		0.79		0.76		0.79	

* Includes the number of unique items; some Biology and English items appear on both Form 108 and Form 208.

Table 10.11 Distributions of Point-Biserial Correlations: MD Mod-HSA May 2008 Operational Items—Linking Sample

	Number and Percentage of Items							
	Algebra		Biology		English		Government	
Correlation	N	%	N	%	N	%	N	%
$R < 0.10$	0	0.0	1	1.0	0	0.0	0	0.0
$0.10 \leq R < 0.20$	4	4.0	5	5.1	0	0.0	1	1.0
$0.20 \leq R < 0.30$	16	16.0	17	17.4	23	23.7	6	6.0
$0.30 \leq R < 0.40$	40	40.0	34	34.7	43	44.3	29	29.0
$0.40 \leq R < 0.50$	35	35.0	36	36.7	28	28.9	52	52.0
$0.50 \leq R < 0.60$	4	4.0	5	5.1	3	3.1	12	12.0
$0.60 \leq R < 0.70$	1	1.0	0	0.0	0	0.0	0	0.0
$R \geq 0.70$	0	0.0	0	0.0	0	0.0	0	0.0
Descriptive Statistics								
N Items*	100		98		97		100	
Mean	0.37		0.37		0.37		0.41	
SD	0.09		0.09		0.08		0.08	
Min	0.18		0.09		0.23		0.13	
Max	0.61		0.54		0.53		0.56	

* Includes the number of unique items; some Biology and English items appear on both Form 108 and Form 208.

Table 10.12 Distributions of Point-Biserial Correlations: MD Mod-HSA May 2008 Operational Items—Target Population

	Number and Percentage of Items							
Correlation	Algebra		Biology		English		Government	
	N	%	N	%	N	%	N	%
$R < 0.10$	1	1.0	6	6.1	3	3.1	2	2.0
$0.10 \leq R < 0.20$	15	15.0	17	17.4	15	15.5	12	12.0
$0.20 \leq R < 0.30$	37	37.0	36	36.7	41	42.3	40	40.0
$0.30 \leq R < 0.40$	37	37.0	33	33.7	32	33.0	41	41.0
$0.40 \leq R < 0.50$	10	10.0	6	6.1	6	6.2	4	4.0
$0.50 \leq R < 0.60$	0	0.0	0	0.0	0	0.0	1	1.0
$0.60 \leq R < 0.70$	0	0.0	0	0.0	0	0.0	0	0.0
$R \geq 0.70$	0	0.0	0	0.0	0	0.0	0	0.0
Descriptive Statistics								
N Items*	100		98		97		100	
Mean	0.29		0.26		0.27		0.29	
SD	0.09		0.10		0.09		0.09	
Min	0.06		-0.03		-0.04		0.05	
Max	0.48		0.48		0.45		0.50	

* Includes the number of unique items; some Biology and English items appear on both Form 108 and Form 208.

Table 10.13 Summary Statistics of Tetrachoric Correlations: MD Mod-HSA May 2008 by Sample, Content, and Form

Sample	Content	Form	Mean	Std Dev	Minimum	Maximum
Linking	Algebra	108	0.206	0.101	-0.120	0.747
		208	0.220	0.102	-0.077	0.581
	Biology	108	0.211	0.106	-0.076	0.524
		208	0.203	0.111	-0.072	0.533
	English	108	0.233	0.094	-0.042	0.627
		208	0.223	0.103	-0.060	0.638
	Government	108	0.266	0.107	-0.072	0.703
		208	0.286	0.104	0.020	0.633
Target	Algebra	108	0.104	0.083	-0.236	0.395
		208	0.099	0.079	-0.114	0.424
	Biology	108	0.071	0.082	-0.196	0.338
		208	0.081	0.088	-0.259	0.402
	English	108	0.081	0.071	-0.180	0.337
		208	0.102	0.077	-0.094	0.346
	Government	108	0.097	0.075	-0.088	0.506
		208	0.097	0.083	-0.145	0.338

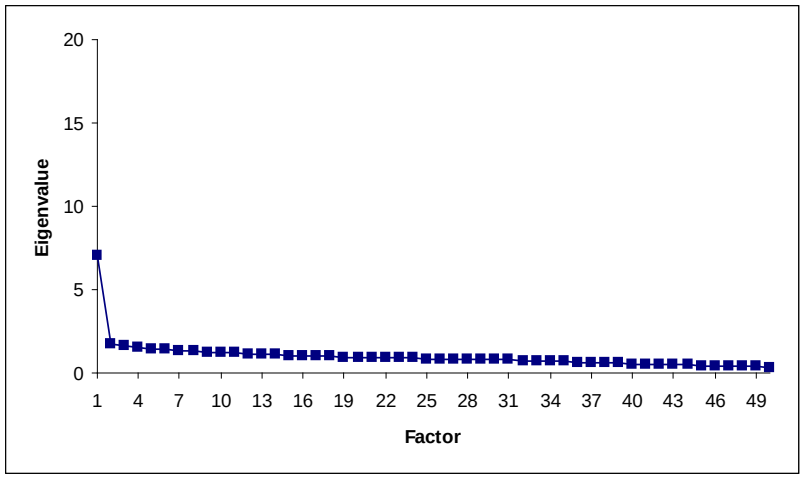

Figure 10.1 Scree Plot: Algebra—Target Population—Form 108

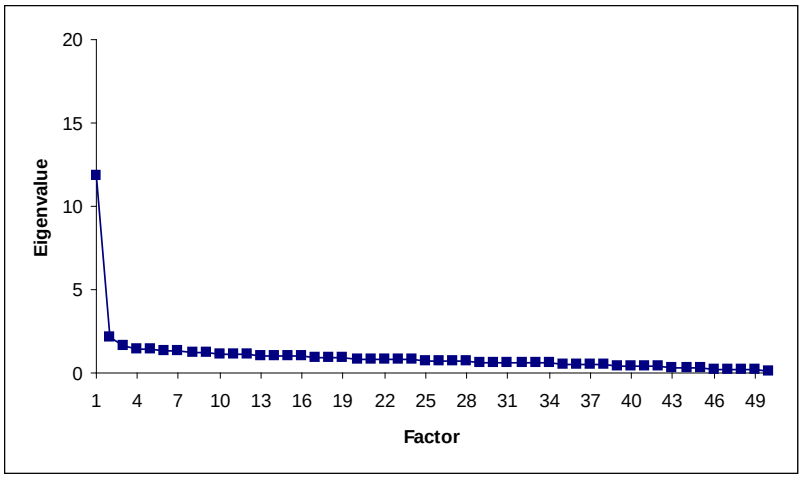

Figure 10.3 Scree Plot: Algebra—Linking Sample—Form 108

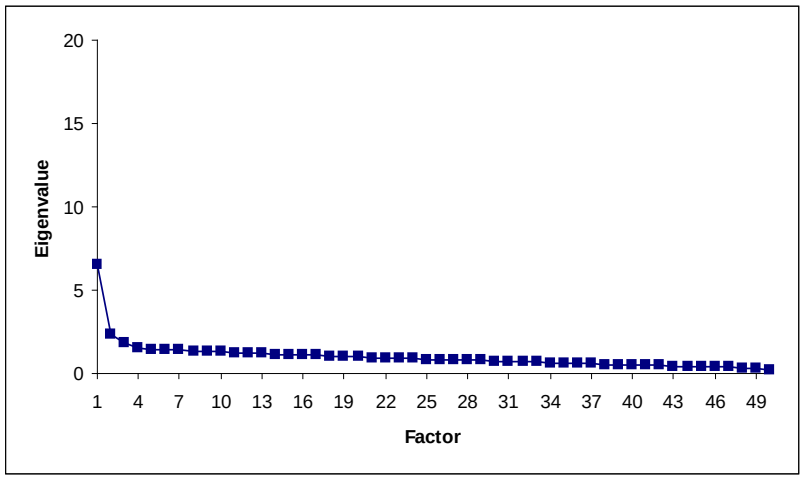

Figure 10.2 Scree Plot: Algebra—Target Population—Form 208

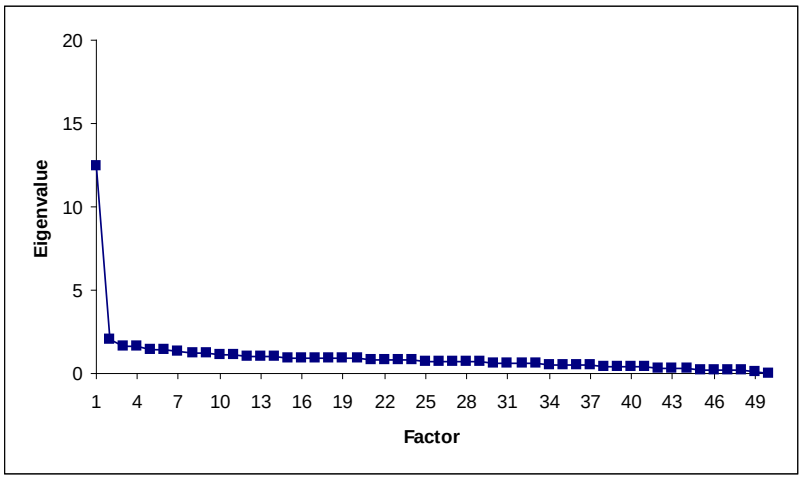

Figure 10.4 Scree Plot: Algebra—Linking Sample—Form 208

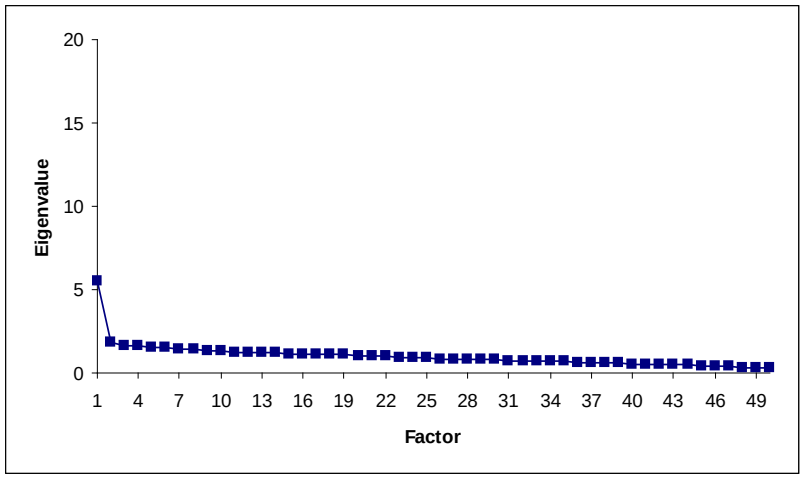

Figure 10.5 Scree Plot: Biology—Target Population—Form 108

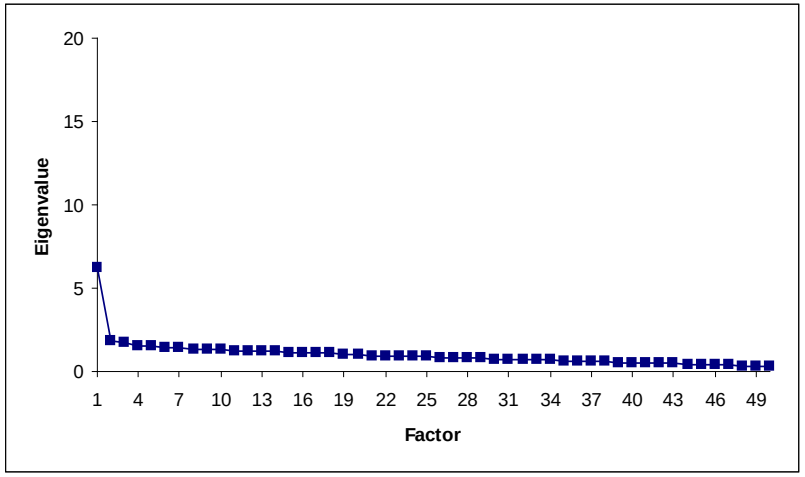

Figure 10.6 Scree Plot: Biology—Target Population—Form 208

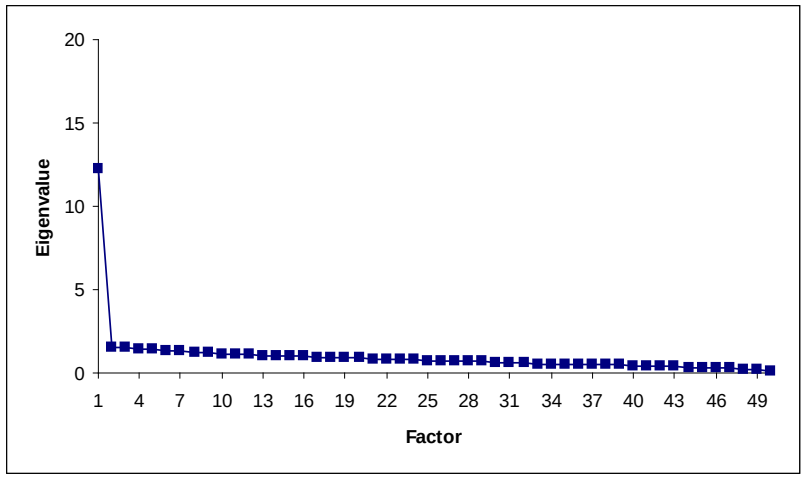

Figure 10.7 Scree Plot: Biology—Linking Sample—Form 108

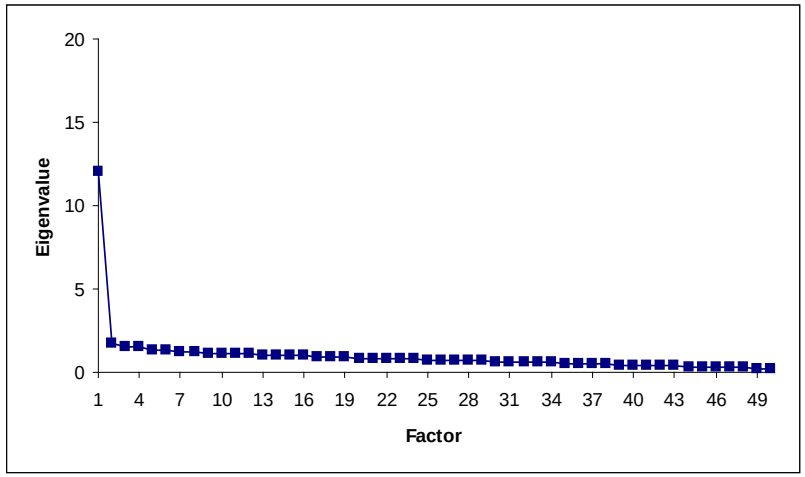

Figure 10.8 Scree Plot: Biology—Linking Sample—Form 208

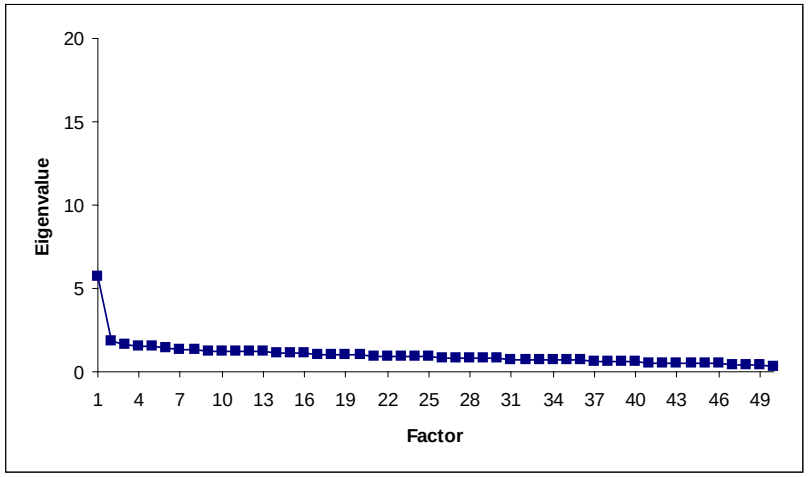

Figure 10.9 Scree Plot: English—Target Population—Form 108

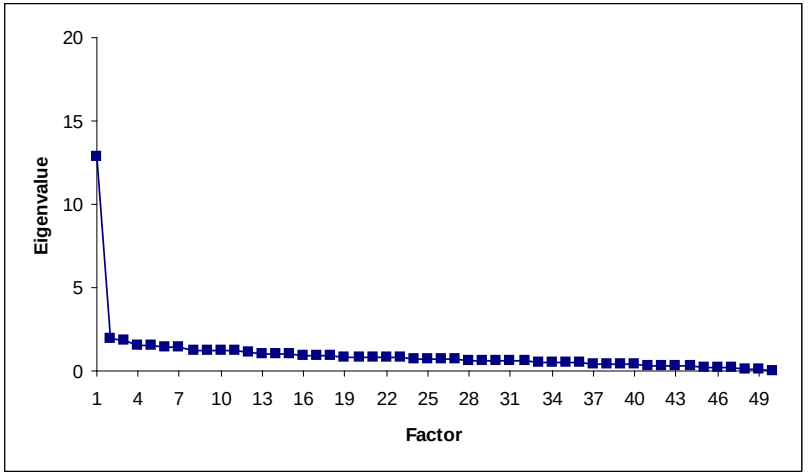

Figure 10.11 Scree Plot: English—Linking Sample—Form 108

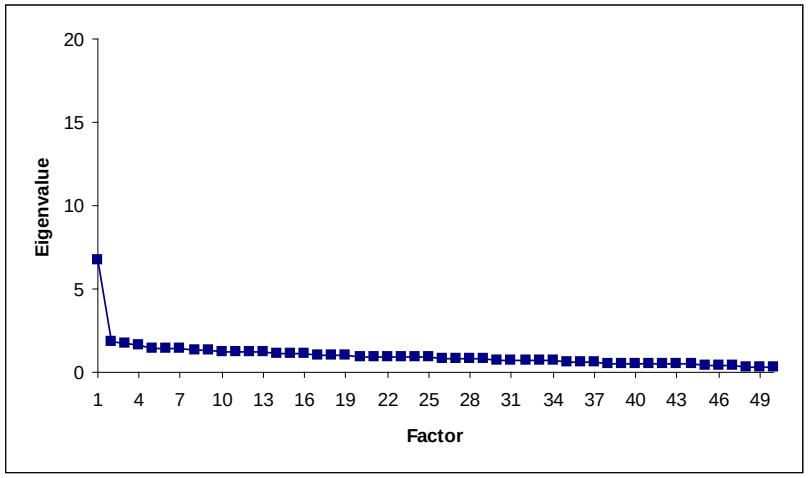

Figure 10.10 Scree Plot: English—Target Population—Form 208

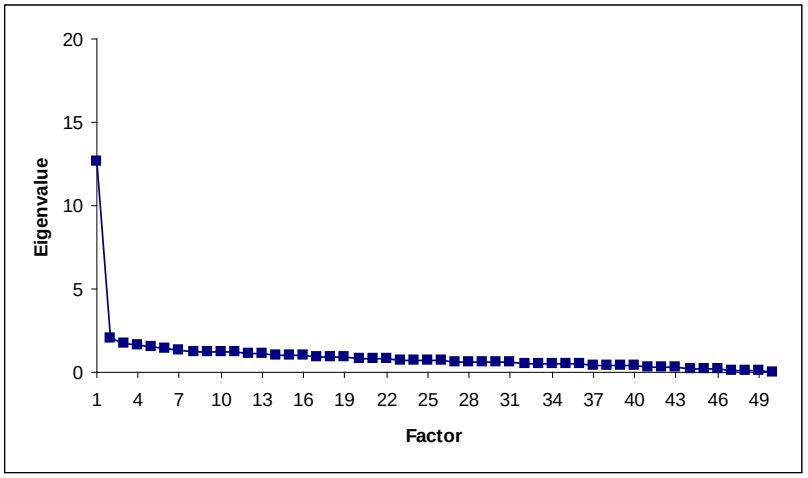

Figure 10.12 Scree Plot: English—Linking Sample—Form 208

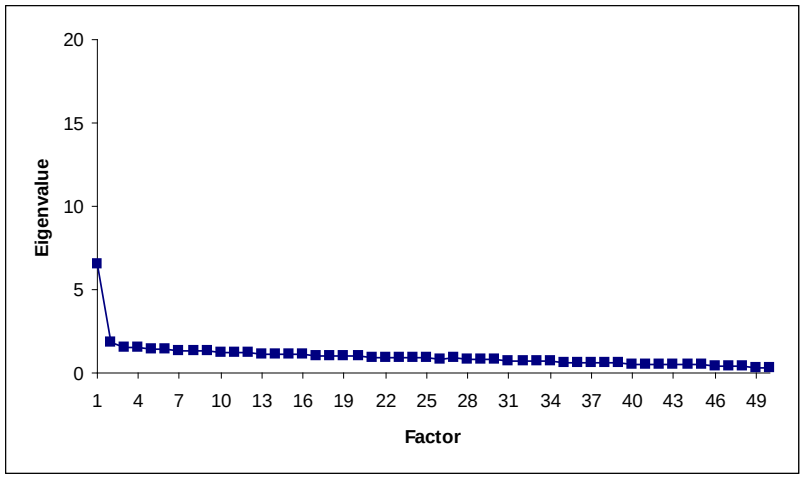

Figure 10.13 Scree Plot: Government—Target Population—Form 108

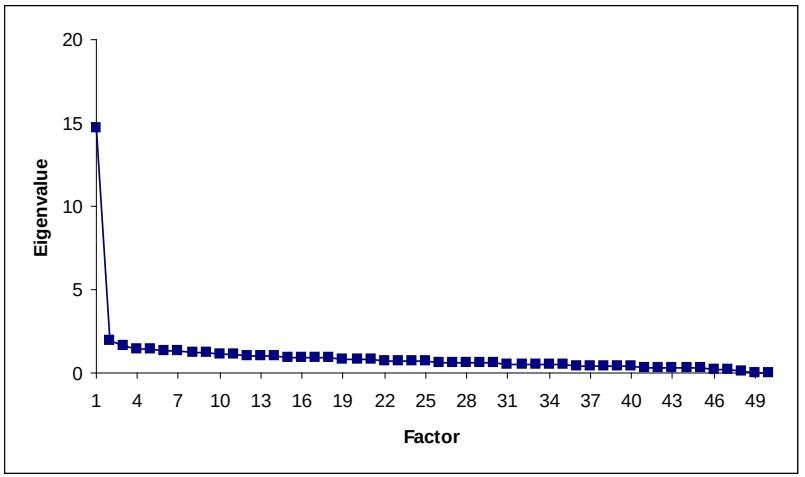

Figure 10.15 Scree Plot: Government—Linking Sample—Form 108

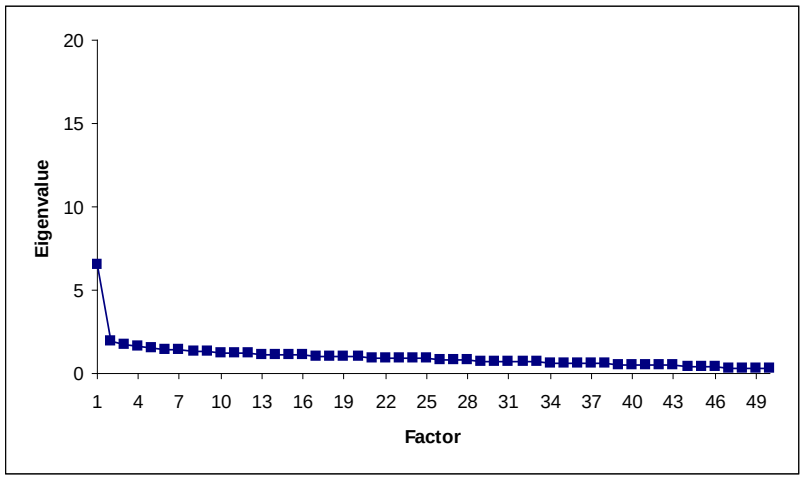

Figure 10.14 Scree Plot: Government—Target Population—Form 208

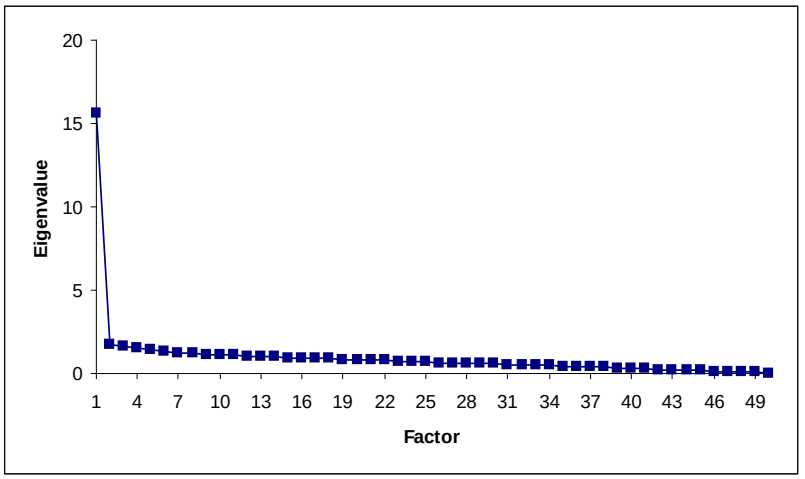

Figure 10.16 Scree Plot: Government—Linking Sample—Form 208

Speededness

The percentage of students who respond to the last items in a test can be used to assess the degree to which a test is speeded. When speededness occurs, a test is measuring not only students' knowledge and skills as defined by the construct of interest but also the speed at which the knowledge and skills are demonstrated, which is a second construct. In tests of achievement, it is desirable to find that speededness is not present in a test, which provides evidence that student scores on the test reflect only the intended construct. Evidence of speededness is provided by the finding that the omit rates at the end of a test are notably higher than those observed elsewhere in the test.

Appendix 2 presents the percentage of students who omitted items on the MD Mod-HSA operational forms. The percentage of students who did not respond to the last ten items of a test was less than 1 percent for all content areas and sessions, with the exception of two forms. The summer administration of Algebra Form P and Government Form Q had 1.6 percent and 1.1 percent of students omitting the last ten items, respectively. These omit rates are comparable to the average omit rates for each form and suggest that students had sufficient time to complete the entire test.

Further, if more than 5 percent of students omit a selected response item at any point in the test, the item is flagged as having a high omit rate. No MD Mod-HSA items were flagged for high omit rate in any content area for any administration.

Other information in support of the uses and interpretations of the MD Mod-HSA scores appears in the following sections:

- Section 11 provides detailed information concerning the scores that were reported and the cut-scores for each content area.
- Section 12 provides information concerning test characteristics based on classical test theory.
- Section 13 presents information regarding student characteristics for the MD Mod-HSA administrations.