TR 2008-02

Validity Evidence Based on Internal Structure: Examination of the Factor Structure of the Ohio Graduation Tests

AIR Technical Team

May 9, 2008

Revision	Description	Revised by	Approval	Date
Int. Rel.	Edited by T. Moore	TWM		5.9.08
TR 2008 – 02				

AMERICAN INSTITUTES FOR RESEARCH

Memorandum

To: Ohio Department of Education

From: AIR Technical Team

Date: 19 February 2008

Re: Validity Evidence Based on Internal Structure: Examination of the Factor Structure of the Ohio Graduation Tests

---

Executive Summary

As part of its review of statewide assessment systems under the No Child Left Behind legislation, the U.S. Department of Education requires states to provide evidence bearing on the validity or appropriateness of interpretations of test scores on statewide achievement tests. In preparation for the U.S. Department of Education review of Ohio's statewide assessment system, the Ohio Department of Education is developing a conceptual framework for demonstrating the validity of test score interpretations resulting from its system of achievement and graduation tests. As the primary contractor for the Ohio Graduation Tests, the American Institutes for Research (AIR) is assisting the Department in collecting and analyzing data addressing the conditions under which, and the degree to which, test scores resulting from the Ohio Graduation Tests can be meaningfully interpreted.

This report describes the results of analyses designed to investigate the internal structure of the Ohio Graduation Tests across subjects and grade levels. In addition to defining the curricular content that Ohio students are expected to achieve, the Ohio Academic Content Standards define a structure of academic achievement in each of the core academic subject areas. Reading achievement is defined by four content standards: Acquisition of Vocabulary, Reading Process, Reading Applications: Literary Text, and Reading Applications: Informational, Technical and Persuasive Text. Student skills in these four content standards are posited to underlie academic

achievement in reading more generally. For mathematics, student skills in the content standards of Data Analysis and Probability; Geometry; Measurement; Number, Number Sense and Operations; and Patterns, Functions and Algebra are posited to underlie student academic achievement in mathematics generally, with high-achieving students demonstrating proficiency across content standards. Writing achievement is defined by three content standards, Writing Applications, Writing Conventions, and Writing Process.

Science achievement is defined by four content standards: Earth and Space Sciences, Life Sciences, Physical Sciences, and Scientific Processes. The Scientific Processes achievement standard is based on the Scientific Inquiry, Scientific Ways of Knowing, and Science and Technology content standards; these content standards were combined in order to ensure reliability of the content standards while maintaining an adequate number of content standards.

Finally for social studies, four content standards, Economics, Government, and Citizenship; People in Societies and Geography; History; and Social Studies Skills and Methods define achievement. As in science, some original social studies content standards were combined into aggregate standards. Three standards – Economics, Government, and Citizenship Rights and Responsibilities – were combined to form the single standard of Economics, Government, and Citizenship; similarly, the Geography and People in Societies content standards were combined to create the People in Society and Geography.

Results reported here are based on the analysis of student responses from the spring 2007 operational administration of the Ohio Graduation Tests for reading, writing, mathematics, science and social studies. Approximately 195,000 students participated in the spring 2007 Ohio Graduation Test administration. The reading, mathematics, science, and social studies achievement tests contain 38 operational items and both multiple-choice and open-ended (constructed response) question formats. The writing achievement test consists of 10 multiple-choice and five extended answer constructed-response items.

Structural equation modeling techniques were employed to perform second-order confirmatory factor analyses on the structure of the Ohio Graduation Tests. Within each subject area, items measure a single standard. Standards within each subject area are, in turn, indicators of

achievement in the subject area. For example, the Reading Test is designed to measure four standards: Acquisition of Vocabulary, Reading Process, Reading Applications: Literary Text, and Reading Applications: Informational, Technical and Persuasive Text. These first-order factors in turn define overall achievement in reading. In these models, each item is an indicator of a single primary factor or academic content standard. In turn, each academic content standard serves as an indicator of achievement in the subject area. Paths from the content standards to the items represent the first-order factor loadings, or the degree to which the items are correlated with the underlying academic content standard construct. Paths from the overall latent variable, reading, to the content standards represent the second-order factor loadings, indicating the degree to which the academic content standard constructs are correlated with the underlying construct of reading achievement.

Although items within a subject area are intercorrelated, each item on the test serves as an indicator for primarily one content standard, and items defining each content standard generally loaded moderately strongly on the intended content standard factor. In turn, the content standards defining the broad subject areas of each achievement domain loaded extremely highly on the intended subject area achievement factor. For reading, mathematics, science and social studies, the structural models defined by the Ohio Academic Content Standards fit the data well. The writing test, however, exhibited reasonable fit but did not meet the fit criteria used to evaluate the model fit for the Ohio Graduation Tests (Hu and Bentler, 1999).

We conclude that the structure of the Ohio Graduation Tests in reading, writing, mathematics, science and social studies conforms to the structural model as defined in the Ohio Academic Content Standards. These results support the validity of the structural model for both the graduation tests and the overall standards framework.

Introduction

As part of its review of statewide assessment systems under the No Child Left Behind legislation, the U.S. Department of Education requires states to provide evidence bearing on the validity or appropriateness of interpretations of test scores on statewide achievement tests. In preparation for the U.S. Department of Education review of Ohio's statewide assessment system, the Ohio Department of Education (ODE) is developing a conceptual framework for demonstrating the validity of test score interpretations resulting from its system of achievement and graduation tests. As the primary contractor for the Ohio Graduation Tests, the American Institutes for Research (AIR) is assisting the Department in collecting and analyzing data addressing the conditions under which, and the degree to which, test scores resulting from the Ohio Graduation Tests can be meaningfully interpreted.

Validity is not an attribute of tests but rather refers to a process whereby the meaning and interpretation of test scores can be established for a variety of intended and even unintended purposes. As such, a test is not valid or invalid; rather, the soundness of interpretations based on test scores is evaluated on the basis of evidence about test properties and test performance across subgroups within the intended (or even unintended) test population.

As the *Standards for Educational and Psychological Testing* (AERA, APA, & NCME, 1999) notes, the validity of any intended score interpretation relies, at its core, on all the evidence accrued about the technical quality of a testing system, including test development and construction procedures, test score reliability, accurate scaling and equating, procedures for setting meaningful performance standards, standardized test administration and scoring procedures, and attention to fairness for all test takers. Evidence for these attributes of the Ohio Graduation Tests are provided as part of the series of technical reports submitted to ODE following each test administration cycle.

Although necessary, these documents, and the evidence they provide, are not sufficient evidence of the validity of Ohio Graduation Test score interpretations. Accrual of validity evidence is an ongoing process, and evidence for the validity of test score interpretation can come from a variety of sources. The *Standards for Educational and Psychological Testing* (AERA, APA, &

NCME, 1999) lists the following sources for validity evidence: evidence based on test content, evidence based on response processes, evidence based on internal structure, evidence based on relations to other variables, and evidence based on consequences of testing.

This document is one of several reports commissioned by the Ohio Department of Education to obtain further evidence about the validity of the achievement tests beyond what is reported in the standard technical reports. This report examines evidence based on the internal structure of the Ohio Graduation Tests administered in March 2007, allowing ODE to evaluate the degree to which relationships among test items and test subscales correspond to the conceptual framework outlined in the Ohio Academic Content Standards.

Ohio has developed a set of academic content standards that define what students should know and be able to do by the end of each grade in each of the major academic subject areas: English language arts, mathematics, science, and social studies. These areas are also to be assessed through the Ohio Graduation Tests as part of Ohio's accountability system under No Child Left Behind. The Academic Content Standards are organized into a three-level framework that includes standards, benchmarks and indicators. Standards represent broad statements about the major content areas within each subject and the learning goals for Ohio students across the K–12 curriculum. Benchmarks denote more specifically what students should know and be able to do at specific points during their schooling. Benchmarks are used to measure student progress toward meeting academic standards. Indicators are specific statements that identify exemplar knowledge and skills that indicate whether students have achieved a benchmark.

The Ohio Graduation Tests measure student progress toward achievement of the Ohio Academic Content Standards. The graduation tests are grade-appropriate, meaning that they measure student progress toward grade-level benchmarks within each academic content standard. The validity of Ohio Graduation Test score interpretations is evaluated along several dimensions. As a criterion-referenced system of tests, the meaning of Ohio Graduation Test scores is, in part, appropriately evaluated by the degree to which test content is aligned with the Ohio Academic Content Standards.

Alignment of content standards is achieved through a rigorous test development process that

stems from the Ohio Academic Content Standards and refers back to those standards in a highly iterative test development process that includes the Ohio Department of Education, the American Institutes for Research (AIR; the test contractor) and two committees composed of Ohio citizens, including not only primary and secondary educators and administrators but also representatives from colleges and universities and other stakeholder groups. The Fairness/Sensitivity Committee (FSC) is charged with ensuring that test content does not unfairly disadvantage students because of group membership, and the Content Advisory Committee (CAC) is specifically charged with ensuring that test content is aligned with academic content standards and is grade-appropriate. Committees review items twice during the item development process, once prior to the inclusion of items on field-test forms and again following field-testing, with item performance information, prior to the inclusion of items in the operational item bank.

This review-intensive item development process ensures that items included in the operational test bank and subsequently appearing on operational test forms are tightly aligned to the Ohio Academic Content Standards and that test scores provide valid feedback to parents and the educational community about student, classroom, school, district, and statewide progress toward achieving those standards.

In addition to defining the content domain for academic achievement in each target subject area, the Ohio Academic Content Standards provide a structural model for student achievement in each subject area. For example, the Ohio Academic Content Standards for English Language Arts (ELA) comprises 10 learning standards:

1. Phonemic Awareness, Word Recognition and Fluency;
2. Acquisition of Vocabulary;
3. Reading Process: Concepts of Print, Comprehension Strategies and Self-Monitoring Strategies;
4. Reading Applications: Informational, Technical and Persuasive Text;
5. Reading Applications: Literary Text;
6. Writing Process;
7. Writing Applications;
8. Writing Conventions;

9. Research;
10. Communication: Oral and Visual.

The first five ELA standards make up the reading subject area, and the three writing standards define the writing subject area. Although the Ohio Graduation Tests are designed to measure student progress toward achievement of standards in both of these subject areas, not all academic content standards are measured in each subject area test. The Phonemic Awareness standard, for example, is primarily targeted to the development of basic skills underlying reading acquisition, including learning to recognize and decode printed words. Students are expected to achieve this standard by the early elementary school years. Thus, although the Phonemic Awareness standard is assessed in the Ohio Diagnostic Tests administered to K–2 students, this standard is not assessed in the Reading Achievement Tests.

In the writing subject area, the emphasis on the Writing Process standard, which focuses on student achievement across each of the major phases of writing, from organizing to drafting and revising text, increases across the school years.

The Ohio Academic Standards for Mathematics have five content standards:

1. Number, Number Sense and Operations;
2. Measurement;
3. Geometry and Spatial Sense;
4. Patterns, Functions and Algebra;
5. Data Analysis and Probability.

There is also a process standard called Mathematical Processes. Each of the five content standards is assessed on each grade-level mathematics achievement test. The Mathematical Processes standard is embedded within each of the content standards and is not independently assessed in the achievement tests.

The Ohio Academic Standards for Science have four content standards:

1. Earth and Space Sciences;
2. Life Sciences;
3. Physical Sciences;

4. Scientific Processes
    a. Scientific Inquiry
    b. Scientific Ways of Knowing
    c. Science and Technology

The Ohio Academic Standards for Social Studies have four content standards:

1. Economics, Government, and Citizenship
    a. Economics;
    b. Government;
    c. Citizenship Rights and Responsibilities;
2. People in Societies and Geography
    a. People in Societies;
    b. Geography;
3. History;
4. Social Studies Skills and Methods.

Methods

Participants. The sample consists of all students who participated in the Ohio Graduation Tests during the March 2007 administration period. Demographic characteristics of the sample are presented in Tables 1 and 2: males made up about 50% of the sample; approximately 75% of the students were White, with African-Americans making up about 19% of the sample, Hispanics about 2%, and Asian/Pacific Islanders about 1%.

Table 1. Number of Achievement Test Respondents by Gender

	Gender				
Grade	Female		Male		Total
	n	%	n	%	
10	76059	49.57	77387	50.43	153446
11	14582	51.48	13745	48.52	28327
12	6751	53.7	5821	46.3	12572
13	42	43.75	54	56.25	96
14	118	69.41	52	30.59	170
TOTAL	97552	50.13	97059	49.87	194611

Table 2. Number of Achievement Test Respondents by Ethnicity

	Ethnicity												
Grade	American Indian		Asian/Pacific Islander		Black/African American		Hispanic		White		Multi-Ethnic		Total
	n	%	n	%	n	%	n	%	n	%	n	%	
10	255	0.2	2217	1.5	22605	14.8	3023	2.0	120712	79.1	2982	2.0	152589
11	52	0.2	321	1.1	8979	31.7	917	3.2	17328	61.2	574	2.0	28301
12	22	0.2	172	1.4	4863	38.7	427	3.4	6802	54.1	227	1.8	12575
13	1	1.0	2	2.0	48	49.0	7	7.1	35	35.7	5	5.1	98
14	1	0.6	1	0.6	64	37.7	9	5.3	82	48.2	5	2.9	170
TOTAL	331	0.2	2713	1.4	36559	18.9	4383	2.3	144959	74.8	3793	2.0	193733

Design. Results are based on the analysis of operational assessments of the Ohio Graduation Tests for reading, writing, mathematics, science and social studies. The Reading, Mathematics, Science and Social Studies Tests contain 38 operational items with 32 multiple-choice and 6 open-ended question formats. The Writing Test contains 13 operational items with 10 multiple-choice items and 3 open-ended items - two writing items provide two scores according to different scoring rubrics, resulting in scores for 5 open-ended items. Structure of the assessments is summarized in Table 3.

Table 3: Design of the Ohio Graduation Tests

Subject	MC items/CR items
Reading	32/6
Writing	10/5
Mathematics	32/6
Science	32/6
Social Studies	32/6

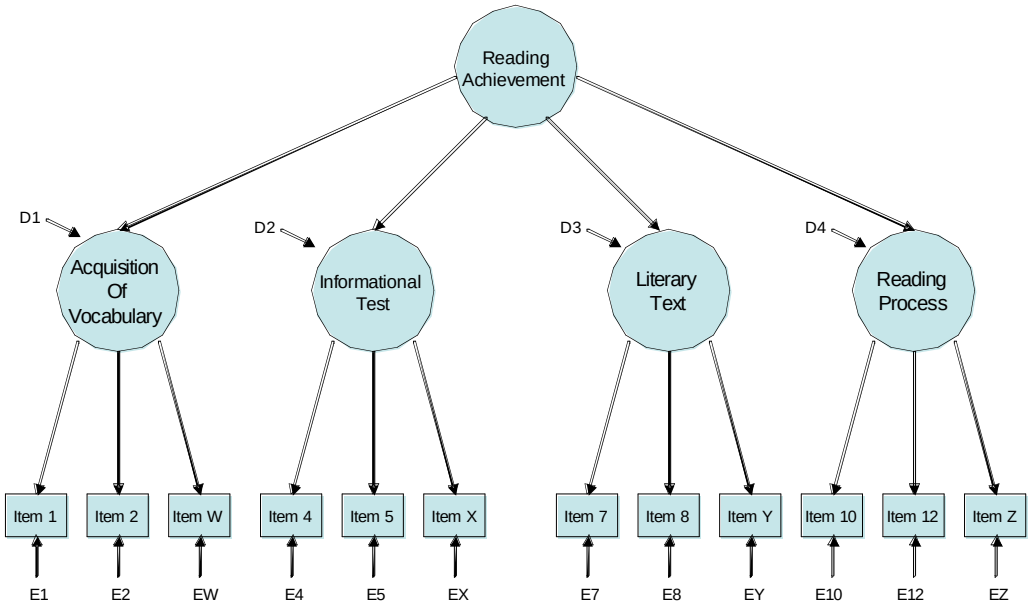
Analyses. AIR used MPlus 4.21 (Muthen & Muthen, 2007), commercially available software for structural equation modeling, to perform second-order confirmatory factor analyses on the structure of the Ohio Graduation Tests. Within each subject area, items measure a single standard. Standards within each subject area are, in turn, indicators of achievement in the subject area.

As noted above, the Reading Tests measure four standards: Acquisition of Vocabulary; Reading Applications: Literary Text; Reading Applications: Informational, Technical and Persuasive Text; and Reading Process. These first-order factors in turn define overall achievement in reading. The form of the second-order confirmatory factor analyses for reading across all grades is illustrated in Figure 1.

As the figure illustrates, each item is an indicator of a single primary factor or academic content standard. Because items are never pure indicators of an underlying factor, each item also includes an error component (denoted as “E” in the figure). Similarly, each academic content standard serves as an indicator of reading achievement. As at the item level, the content standards include an error term (denoted as “D” in the figure) indicating that the content

standards are not pure indicators of overall reading achievement. The paths from the content standards to the items represent the first-order factor loadings, the degree to which items are correlated with the underlying academic content standard construct. Similarly, the paths from reading achievement to the content standards represent the second-order factor loading, indicating the degree to which academic content standard constructs are correlated with the underlying construct of reading achievement.

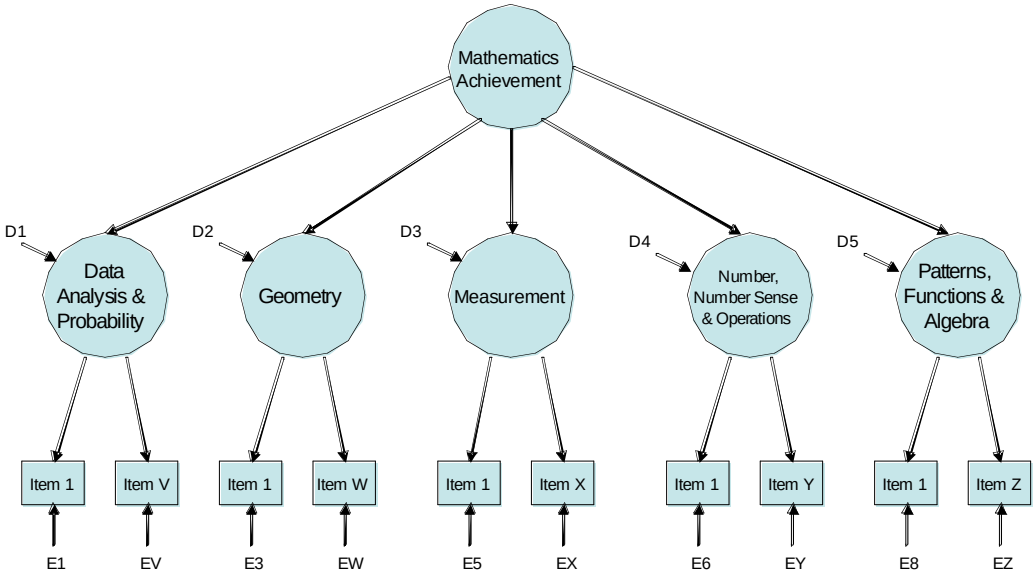
Figure 1. Second-Order Confirmatory Factor Analysis Model for Reading Achievement*



*Note. The number of items per factor shown in the figure is for illustrative purposes only. In the 2007 Reading Graduation Test, the number of items per factor ranged from 7 to 12, depending on content standard.

The Mathematics Graduation Test measures five standards: Data Analysis and Probability; Geometry; Measurement; Number, Number Sense and Operations; and Patterns, Functions and Algebra. These first-order factors in turn define overall achievement in mathematics. The form of the second-order confirmatory factor analyses for mathematics across all grades is illustrated in Figure 2.

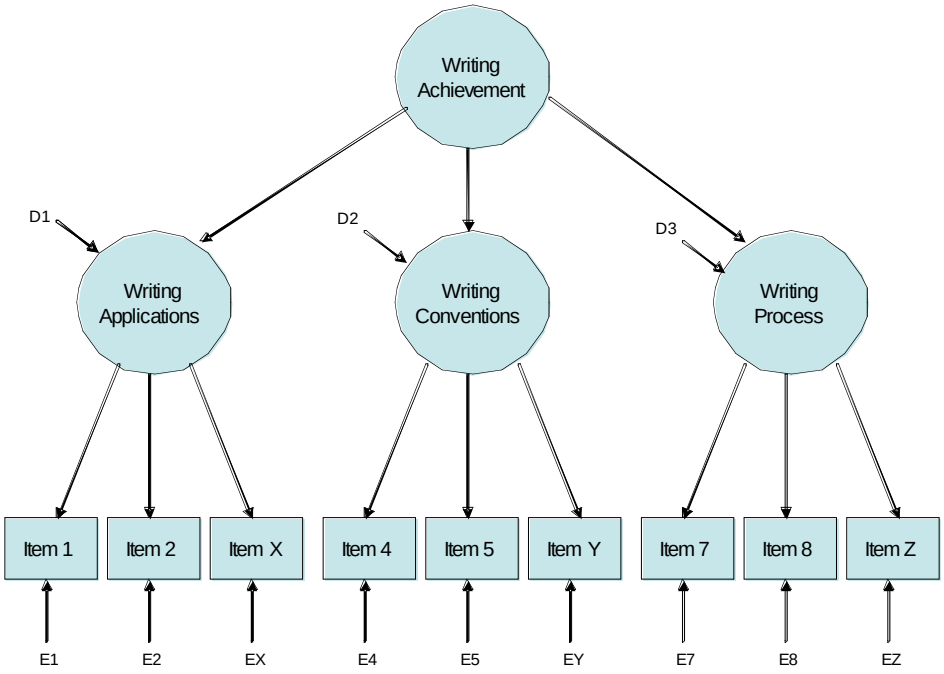
Figure 2. Second-Order Confirmatory Factor Analysis Model for Mathematics Achievement



*Note. The number of items per factor shown in the figure is for illustrative purposes only. In the 2007 Mathematics Graduation Test, the number of items per factor ranged from 6 to 10, depending on content standard.

The Writing Graduation Test measures three standards: Writing Applications, Writing Conventions, and Writing Process. These first-order factors in turn define overall achievement in writing. The form of the second-order confirmatory factor analysis for writing is illustrated in Figure 3.

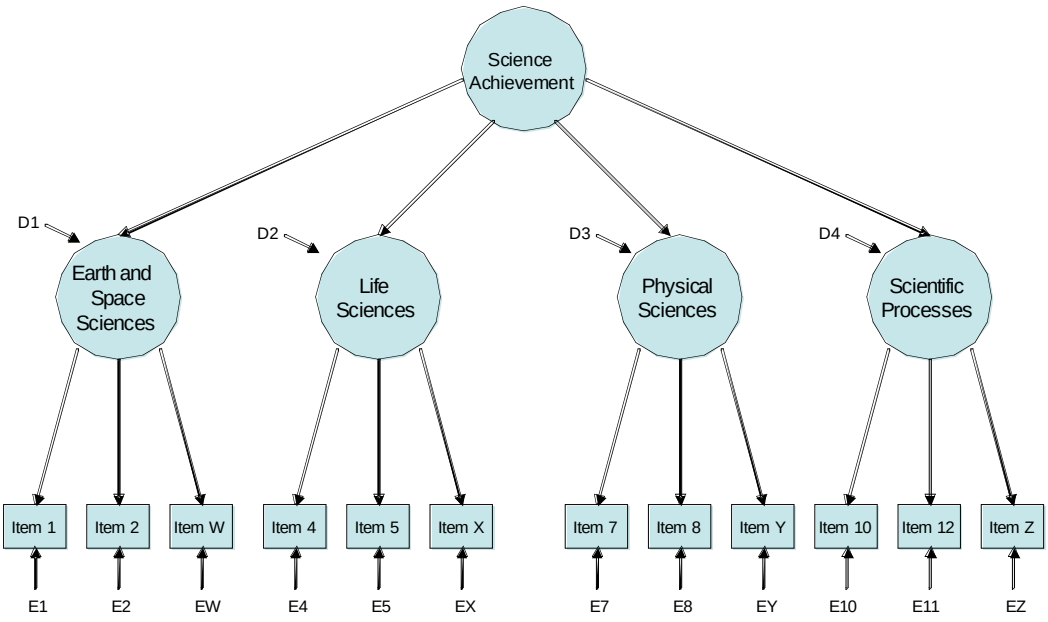
Figure 3. Second-Order Confirmatory Factor Analysis Model for Writing Achievement



*Note. The number of items per factor shown in the figure is for illustrative purposes only. In the 2007 Writing Graduation Test, two items contributed to the assessment of the Writing Applications and Writing Conventions standards and 11 items assessed the Writing Process standard.

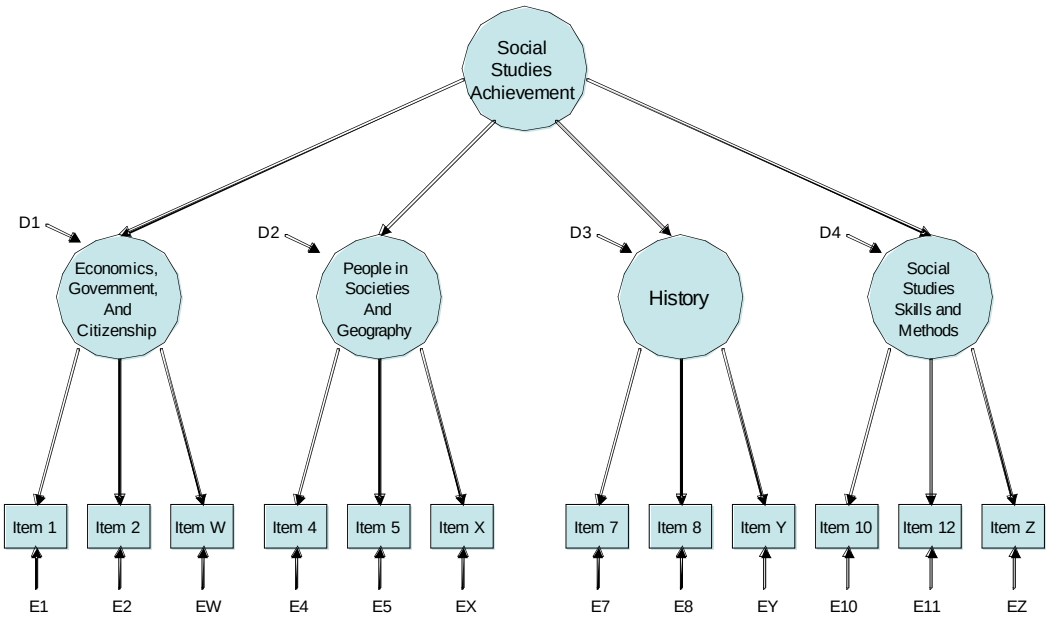
The Science Graduation Test measures four standards: Earth and Space Sciences, Life Sciences, Physical Sciences, and Scientific Processes. The Social Studies Graduation Test measures four standards: Economics, Government, and Citizenship; People in Societies and Geography; History; and Social Studies Skills and Methods. These first-order factors in turn define overall achievement on science and social studies achievement tests. The factor structure of science and social studies graduation tests are shown in Figures 4 and 5.

Figure 4. Second-Order Confirmatory Factor Analysis Model for Science Achievement



*Note. The number of items per factor shown in the figure is for illustrative purposes only. In the 2007 Science Graduation Test, the number of items per factor ranged from 9 to 11, depending on content standard.

Figure 5. Second-Order Confirmatory Factor Analysis Model for Social Studies Achievement



*Note. The number of items per factor shown in the figure is for illustrative purposes only. In the 2007 Social Studies Graduation Test, the number of items per factor ranged from 9 to 11, depending on content standard and grade level.

Factor Analytic Methods: A series of confirmatory factor analyses (CFA) were conducted using the statistical program Mplus for each grade and subject assessment. These psychometric tools are useful data analytic methods commonly used for collecting validity evidence on the internal structure of assessments.

In factor analytic terms, a test structure that is strictly unidimensional implies a single-order factor model where all test items load onto a single underlying factor. This model is subsequently referred to as the *base model* and is used as the basis for comparison. If the internal structure of the test were strictly unidimensional, then the overall person ability measure, theta (θ), would be the single common factor and the correlation matrix among test items would suggest no

discernable pattern among factors. As such, there would be no empirical or logical basis to report scores for the separate performance categories.

As previously stated, the method of reporting scores used for the Ohio Graduation Test implies separate factors for each reporting category connected by a single underlying factor. This model is subsequently referred to as the *implied model.* In factor analytic terms, this suggests that test items load onto separate first-order factors with the first-order factors connected to a single underlying second-order factor.

Fortunately, empirical methods such as confirmatory factor analysis (CFA) exist thus making it possible to collect validity evidence on the degree to which the implied model is reasonable when compared to the base model. Hence, the purpose of the subsequent analyses is to collect validity evidence on the measurement model and reporting structures used for the Ohio Graduation Tests via a direct comparison of the implied and base models.

The factor analysis models are based on the matrix $\mathbf{S}$ of tetrachoric and polychoric sample correlations among the item scores (Olsson, 1979) and the matrix $\mathbf{W}$ of asymptotic covariances among these sample correlations (Jöreskog, 1994), which is employed as the weighting matrix in a weighted least squares estimation approach (Browne, 1984; Muthén, 1984) to minimize the fit function:

$$F_{WLS} = \text{vech}(\mathbf{S} - \hat{\Sigma})' \mathbf{W}^{-1} \text{vech}(\mathbf{S} - \hat{\Sigma}) \qquad (0.1)$$

In (0.1), $\hat{\Sigma}$ is the implied correlation matrix, given the estimated factor model, and the function *vech* vectorizes a symmetric matrix. That is, *vech* stacks each column of the matrix to form a vector.

Testing the implied model against the base model requires development of two factor analyses for each grade and subject. For the base model, a first-order factor analysis was posited where all test items load onto a single common factor. The first-order model can be mathematically represented as:

$$\hat{\Sigma} = \Lambda\Phi\Lambda' + \Theta \qquad (0.2)$$

where Λ is the matrix of item factor loadings (with Λ' representing its transpose), and Φ is the

uniqueness, or measurement error. The matrix $\div$ is the correlation among the separate factors. For the base model, items are thought only to load onto a single underlying factor. Hence $\mathbf{h}$ is a *p x 1* vector where *p* is the number of test items and $\div$ is a scalar equal to 1. Therefore, it is possible to drop the matrix $\div$ from the general notation. However, this notation is retained to more easily facilitate comparisons with the implied model, such that the base model can subsequently be viewed as a special case of the second-order factor analysis.

For the implied model, a second-order factor analysis is posited where test items are coerced to load onto the reporting categories they are designed to target and all reporting categories share a common underlying factor. The second-order factor analysis can be mathematically represented as:

$$\mathbf{H} = \mathbf{h}(\sqrt{}\div\sqrt{}'+\mathbf{y})\mathbf{h}'+\mathbf{»} \qquad (0.3)$$

where $\mathbf{H}$ is the implied correlation matrix among test items, $\mathbf{h}$ is the *p x k* matrix of first-order factor loadings relating item scores to first-order factors, $\sqrt{}$ is the *k x 1* matrix of second-order factor loadings relating the first-order factors to the second-order factor with *k* denoting the number of factors, $\div$ is the correlation matrix of the second-order factors , and $\mathbf{y}$ is the matrix of first-order factor residuals. All other notation is the same as the first-order model. Note that the second-order model expands the first-order model such that $\div \rightarrow \sqrt{}\div\sqrt{}'+\mathbf{y}$. As such, the first-order model is said to be nested within the second-order model.

Results

Factor loadings for each subject are presented in the appendix; in each table, item-loading on its intended scale is shown for each of the first-order factors, followed by the loadings for each standard on the second-order factor. Loadings of the one-factor model are shown on the subject achievement column. Loadings for items are generally moderate to high, loading above 0.3. The primary factor standards typically load above 0.9 on the general, second-order achievement factor.

Joint fit criteria proposed by Hu and Bentler (1999), which account for differences in model specification such as sample size and model complexity, were used to evaluate the model fit:

> Comparative Fit Index (CFI; Bentler, 1990) over .96 and the Standardized Root-mean-squared Residual (SRMR; Bentler, 1995) less than 0.10; or
> SRMR less than 0.10 and the Root Mean Square Error of Approximation (RMSEA; Steiger and Lind, 1980) less than 0.06.

Reading. Table A-1 presents the standardized factor loadings for the reading solution. The loading for items are generally moderate, with one item loading below 0.3 on their intended factor. Item 44 loads 0.44 on the Informational Text standard. All four of the primary factor standards load above .9 on the general, second-order reading achievement factor. The second-order model meets the joint criteria (CFI=0.862, RMSEA=0.021 and SRMR=0.044) indicating reasonable model fit.

Mathematics. Table A-2 presents the standardized factor loadings for the mathematics solution. Although average item loadings are generally moderate in magnitude, two items have loadings below .3 on their intended factor. Item 23 has a loading of 0.173 on the Data Analysis and Probability factor; item 24 loads weakly on the Patterns, Functions, and Algebra first-order factor. All five of the primary factor standards load above 0.9 on the general, second-order mathematics achievement factor. The second-order model meets the joint criteria (CFI=0.929, RMSEA=0.017 and SRMR=0.030) indicating reasonable model fit.

Writing. Table A-3 presents the standardized factor loadings for the writing solution. The Writing Applications and Writing Conventions factors are defined by only two items, where the

same two items provide scores for both factors through different scoring rubrics. Although average item loadings are generally moderate in magnitude, two items have loadings below .3 on their intended factor. Item 2 has a loading of 0.173 and item 12 has a loading of 0.193 both on the Writing Process factor. All of the primary factor standards load above .9 on the second-order writing achievement factor. The model does not meet the joint criteria (CFI=0.951, RMSEA=0.033), although it closely approaches the criteria threshold indicating a high likelihood of reasonable model fit. SRMR cannot be computed for the writing test because it contains two item formats: categorical and continuous. Under Mplus, categorical items can have a maximum of 10 score points; while all other models contain only items specified as categorical variables, writing also contains items with greater than 10 score points - referred to as "continuous" variables.

Science. Standardized factor loadings for the science solution are shown in Table A-4. The loading for items are generally moderate, with no item loading below 0.3 on their intended factor. The four primary factor standards load above 0.9 on the second-order science achievement factor. The model meets the joint criteria (CFI=0.894, RMSEA=0.016 and SRMR=0.029) indicating reasonable model fit.

Social Studies. Standardized factor loadings for the social studies solution are shown in Table A-5. The loading for items are generally moderate, with no item loading below 0.3 on their intended factor. The four primary factor standards load above 0.9 on the second-order science achievement factor. The model meets the joint criteria (CFI=0.920, RMSEA=0.016 and SRMR=0.033) indicating reasonable model fit.

Fit Summary

Tables 4a and 4b summarize model-fit indices for the one factor models and second-order factor models, described in-text above for each subject and grade. Models for reading, mathematics, science, and social studies are shown to meet the joint criteria as specified by Hu and Bentler (1999); each subject produced Standardized Root-Mean-Squared Residuals (SRMR) less than .10 and Root Mean Square Error of Approximation (RMSEA) less than 0.06.
As stated previously, SRMR cannot be computed for the writing test because it contains two item

formats: categorical and continuous. The writing test does not meet the joint criteria because of CFI index score of 0.951, which is slightly smaller that the cut score of 0.96. It is very likely that the writing test has a reasonable model fit.

Table 5 presents the chi-square statistics for each subject representing the difference between the chi-squares obtained from the base, one-factor model and the implied, second-order model. The column *DF* indicates the degrees of freedom – the difference between the number of parameters estimated in the base model and the implied model. The last column presents the critical value from the chi-square distribution corresponding to an alpha of 0.05, given the degrees of freedom in the column *DF*.

Significant results indicate that data support the validity of the second-order model with regards to the validity of the Ohio Graduation Test content standards. The results displayed in Table 5 demonstrate that the second-order model fits significantly better than the one factor model.

Table 4a. Summary of Model Fit Statistics for the One Factor Models

	χ^2	**# of Free Parameters**	**CFI**	**RMSEA**	**SRMR**
Reading	46962.5	38	0.861	0.021	0.044
Math	34136.8	38	0.924	0.017	0.032
Writing	32062.7	17	0.9	0.047	
Science	8797.0	38	0.891	0.016	0.030
Social	31870.5	38	0.918	0.017	0.033

Table 4b. Summary of Model Fit Statistics for the Second-Order Models

	χ^2	**# of Free Parameters**	**CFI**	**RMSEA**	**SRMR**
Reading	46474.8	42	0.862	0.021	0.044
Math	31746.4	43	0.929	0.017	0.030
Writing	15598.1	20	0.951	0.033	
Science	31391.2	42	0.894	0.016	0.029
Social	31229.5	42	0.920	0.016	0.033

Table 5. Summary of Likelihood Ratio Test Between One-Factor and Second-Order Models

	χ^2	**DF**	**Critical Value**
Reading	488	4	9.49
Writing	16465	3	7.82
Math	2390	5	11.07

Science	765	4	9.49
Social	641	4	9.49

Conclusion

This report describes the results of analyses designed to investigate the internal structure of the Ohio Graduation Tests across subjects and grade levels. In addition to defining the curricular content that Ohio students are expected to achieve, the Ohio Academic Content Standards define a structure of academic achievement in each core academic subject area. Reading achievement is defined by four content standards: Acquisition of Vocabulary, Reading Process, Reading Applications: Literary Text, and Reading Applications: Informational, Technical, and Persuasive Text. Student skills in these four content standards are posited to underlie academic achievement in reading more generally. Similarly for writing, mathematics, science, and social studies, student skills in the content standards are posited to underlie student academic achievement in the given subject area generally, with high-achieving students demonstrating proficiency across content standards.

For all subjects, except for writing, the structural models defined by the Ohio Academic Content Standards meet the joint criteria proposed by Hu and Bentler (1999; CFI > 0.96 and SRMR ≤ 0.10 or SRMR ≤ 0.10 and RMSEA ≤ 0.06). The writing test does not meet the joint criteria as the observed CFI statistic (0.951) is below the threshold of 0.96. However, it is very likely that the second order factor model for writing evidences acceptable model given that the CFI value is very close to the threshold and RMSEA (0.033) meets specified criteria (RMSEA ≤ 0.06). Two possible reasons for this slightly lower CFI value: 1) Two of three factors on the writing test are measured by only two items, or 2) there are mixed types of items, categorical and continuous. Although items within a subject area are intercorrelated, each item on the test serves as an indicator for primarily one content standard, and items defining each content standard generally load moderately strongly on the intended content standard factor. In turn, the content standards defining the broad subject areas of reading, mathematics, writing, science and social studies achievement load extremely highly on the intended subject area achievement factor.

In summary, the structure of the Ohio Graduation Tests in reading, mathematics, writing, science and social studies conforms to the associated structural model of achievement as defined in the Ohio Academic Content Standards. These results support the validity of the structural model for both the achievement tests and the overall standards framework.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M. (2002). *EQS 6 structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., and Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 37,* 1-21.

Hu, L., and Bentler, PM. (1999). Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria versus New Alternatives. *Structural Equation modeling, 6*(1), p1-55.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychonetrika, 59,* 381-389.

Muthén, B. O. (1984). A general structure equation model with dichotomous, ordered categorical and continuous latent variable indicators. *Psychometrika, 49,* 115-132.

Muthén, L., and Muthén, B. O. (2007). *Mplus 4.1* [Computer Software]. Los Angeles, CA: Muthén & Muthén.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika, 44*(4), 443-460.

Steiger, J. H., and Lind, J. C. (1980). *Statistically-based tests for the number of common factor*s. Paper presented at the annual Spring Meeting of the Psychometric Society in Iowa City. May 30, 1980.

Appendix: Standardized Factor Loading Tables by Subject and Grade

Table A-1. Standardized Factor Loadings —Reading

	AV	IT	LT	RP	Reading Achievement
ITEM 2	0.551				0.541
ITEM 6	0.617				0.606
ITEM 14	0.333				0.325
ITEM 15	0.347				0.34
ITEM 24	0.739				0.725
ITEM 30	0.372				0.365
ITEM 31	0.533				0.525
ITEM 3		0.378			0.376
ITEM 4		0.346			0.344
ITEM 5		0.371			0.37
ITEM 21		0.753			0.75
ITEM 22		0.683			0.68
ITEM 26		0.69			0.687
ITEM 27		0.791			0.788
ITEM 28		0.706			0.704
ITEM 29		0.721			0.719
ITEM 42		0.524			0.522
ITEM 43		0.349			0.349
ITEM 44		0.185			0.185
ITEM 13			0.499		0.498
ITEM 16			0.466		0.465
ITEM 18			0.703		0.701
ITEM 19			0.687		0.686
ITEM 20			0.588		0.587
ITEM 32			0.757		0.755
ITEM 35			0.706		0.705
ITEM 36			0.769		0.767
ITEM 37			0.301		0.3
ITEM 38			0.316		0.315
ITEM 39			0.6		0.599
ITEM 1				0.715	0.713
ITEM 17				0.815	0.815
ITEM 23				0.368	0.368
ITEM 25				0.668	0.667
ITEM 33				0.46	0.46
ITEM 34				0.373	0.373
ITEM 40				0.302	0.3
ITEM 41				0.669	0.668
AV					0.978
IT					0.989
LT					0.996
RP					1

**Note: AV = Acquisition of Vocabulary; IT = Informational Text; LT = Literary Text; RP = Reading Process.*

Table A-2. Standardized Factor Loadings —Mathematics

	DA	G	M	NN	PF	Mathematics Achievement
ITEM 1	0.459					0.448
ITEM 5	0.389					0.381
ITEM 17	0.782					0.766
ITEM 20	0.477					0.464
ITEM 23	0.173					0.165
ITEM 26	0.628					0.616
ITEM 31	0.635					0.622
ITEM 34	0.662					0.651
ITEM 39	0.639					0.626
ITEM 43	0.606					0.592
ITEM 6		0.443				0.436
ITEM 10		0.746				0.725
ITEM 25		0.779				0.76
ITEM 29		0.382				0.375
ITEM 33		0.47				0.458
ITEM 37		0.73				0.713
ITEM 42		0.594				0.58
ITEM 4			0.732			0.727
ITEM 9			0.704			0.702
ITEM 21			0.641			0.639
ITEM 27			0.563			0.561
ITEM 36			0.664			0.662
ITEM 41			0.615			0.613
ITEM 3				0.727		0.724
ITEM 8				0.712		0.709
ITEM 19				0.674		0.671
ITEM 28				0.831		0.826
ITEM 32				0.679		0.675
ITEM 35				0.497		0.496
ITEM 40				0.656		0.653
ITEM 2					0.473	0.473
ITEM 7					0.708	0.708
ITEM 18					0.565	0.566
ITEM 22					0.814	0.812
ITEM 24					0.189	0.189
ITEM 30					0.581	0.582
ITEM 38					0.837	0.836
ITEM 44					0.448	0.448
DA						0.97
GS						0.966

ME						0.993
NS						0.993
PA						1

**Note: DA = Data Analysis & Probability; G = Geometry; M = Measurement; NN = Number, Number Sense & Operations; PF = Patterns, Functions & Algebra.*

Table A-3. Standardized Factor Loadings —Writing

	Writing Applications	Writing Conventions	Writing Process	Writing Achievement
ITEM 1A	0.83			0.825
ITEM 13A	0.808			0.792
ITEM 1C		0.886		0.854
ITEM 13C		0.871		0.858
ITEM 2			0.233	0.223
ITEM 3			0.467	0.436
ITEM 4			0.44	0.411
ITEM 5			0.624	0.573
ITEM 6			0.464	0.428
ITEM 7			0.704	0.653
ITEM 8			0.351	0.316
ITEM 9			0.573	0.534
ITEM 10			0.599	0.554
ITEM 11			0.602	0.559
ITEM 12			0.193	0.195
Writing Applications				0.904
Writing Process				1
Writing Process				0.894

Table A-4. Standardized Factor Loadings — Science

	ES	LS	PS	SP	Science Achievement
ITEM 6	0.54				0.538
ITEM 18	0.598				0.595
ITEM 19	0.523				0.521
ITEM 24	0.559				0.556
ITEM 29	0.557				0.555
ITEM 31	0.601				0.598
ITEM 32	0.654				0.65
ITEM 35	0.598				0.596
ITEM 38	0.504				0.501
ITEM 2		0.445			0.443
ITEM 3		0.408			0.407
ITEM 4		0.486			0.484
ITEM 5		0.46			0.458
ITEM 25		0.593			0.591
ITEM 26		0.564			0.562
ITEM 27		0.349			0.348
ITEM 30		0.538			0.536
ITEM 40		0.766			0.763
ITEM 42		0.601			0.598
ITEM 44		0.633			0.63
ITEM 1			0.518		0.511
ITEM 7			0.664		0.652
ITEM 8			0.586		0.579
ITEM 9			0.56		0.552
ITEM 10			0.419		0.413
ITEM 20			0.612		0.604
ITEM 21			0.641		0.632
ITEM 22			0.404		0.397
ITEM 23			0.413		0.406
ITEM 36			0.478		0.472
ITEM 41			0.736		0.725
ITEM 37				0.745	0.741
ITEM 39				0.712	0.708
ITEM 43				0.34	0.338
ITEM 17				0.582	0.58
ITEM 33				0.575	0.572
ITEM 28				0.654	0.651
ITEM 34				0.492	0.49
ES					0.994
LS					0.996
PS					0.976
SP					0.991

**Note: ES = Earth and Space Sciences; LS = Life Sciences; PS = Physical Sciences; SP = Scientific Processes.*

Table A-5. Standardized Factor Loadings —Social Studies

	EG	SG	HI	SO	Social Studies Achievement
ITEM 6	0.736				0.735
ITEM 21	0.825				0.823
ITEM 29	0.729				0.727
ITEM 34	0.427				0.426
ITEM 8	0.711				0.71
ITEM 26	0.626				0.625
ITEM 39	0.636				0.635
ITEM 20	0.631				0.63
ITEM 33	0.791				0.79
ITEM 3		0.721			0.72
ITEM 17		0.511			0.511
ITEM 22		0.663			0.663
ITEM 30		0.538			0.538
ITEM 37		0.623			0.623
ITEM 43		0.393			0.393
ITEM 4		0.414			0.414
ITEM 10		0.748			0.747
ITEM 25		0.651			0.65
ITEM 32		0.712			0.712
ITEM 41		0.678			0.677
ITEM 1			0.51		0.506
ITEM 5			0.452		0.448
ITEM 9			0.674		0.67
ITEM 19			0.77		0.765
ITEM 23			0.719		0.713
ITEM 27			0.512		0.508
ITEM 31			0.658		0.655
ITEM 36			0.615		0.612
ITEM 40			0.548		0.544
ITEM 42			0.441		0.438
ITEM 2				0.7	0.692
ITEM 7				0.623	0.617
ITEM 18				0.662	0.656
ITEM 24				0.603	0.597
ITEM 28				0.655	0.649
ITEM 35				0.634	0.627
ITEM 38				0.636	0.629
ITEM 44				0.598	0.592
EG					0.998
SG					1
HI					0.988
SO					0.984

**Note: EG = Economics, Government, and Citizenship; PS = People in Societies and Geography; H = History; SS = Social Studies Skills and Methods.*