Assessment of Science Proficiency in Grade 4: Establishing Valid and Reliable Measures using Mixed Methods

Todd Sherron, Ph.D.
info2knowledge, llc.

Carol Fletcher, Ph.D.
The University of Texas at Austin
Center for Science and Mathematics Education
The Texas Regional Collaboratives

Paper presented at the Association for Science Teacher Education 2007 International Conference, January, 3rd, 2006. Correspondence should be sent Todd Sherron, 4810 Summerset Trail, Austin, TX 78749, or sherron@i2-k.com. The authors would like to thank Lisa Bellows, John Lineweaver, David Allard, Kathy Coker, Thomas Gadsden, and Virginia Malone for their contributions that made this project successful.

ABSTRACT

The grade 4 Science Field Test was constructed to aid science educators as well as science researchers in understanding the relationship between professional development for science teachers and student achievement. Our overarching research agenda is to examine the correlation between Texas Regional Collaboratives (TRC) professional development science teachers' metrics and student achievement. The TRC is a statewide network of P-16 partnerships that provide sustained and high intensity professional development to P-12 teachers of science across the state. For more information about TRC use the link: www.TheTRC.org. The grade 4 Science Field Test is web based and delivers online reports that will help science educators identify strengths and weaknesses in their science program. The science field test provides a correlated diagnostic measure of student achievement in science at the end of the fourth grade that can be used to guide instruction at the 5th grade level. As with any study, the foundation is built upon valid and reliable scores. Herein we overview the methodology used to develop the grade 4 Science Field Test and present the psychometric properties of the test. The grade 4 Science Field Test is correlated to Texas Essential Knowledge and Skills (TEKS) and models the Texas Assessment of Knowledge and Skills (TAKS) test by measuring four constructs: (1) Nature of Science; (2) Life Science; (3) Physical Science; (4) Earth Science.

BACKGROUND

The Texas Regional Collaboratives for Excellence in Science Teaching (TRC) is a statewide network of P-16 partnerships that provide sustained and high intensity professional development to P-12 teachers of science. This infrastructure of over 43 institutions of higher education collaborating with the Texas Education Agency, education service centers, school districts, and business partners has a 15-year track record of designing and implementing science teacher professional development. Funding for TRC partnerships comes from the Department of Education's Math and Science Partnership and yearly goals and objectives are developed in concert with the Texas Education Agency.

In 2005, the State of Texas recorded a 64% pass rate for English Science Texas Assessment of Knowledge and Skills (TAKS). Only 23% of the fifth grade students passed the Spanish Science TAKS. According to Chris Castillo-Comer, Director of Science at the Texas Education Agency, "More elementary students failed the science portion of the TAKS test than any other subject."

PURPOSE

The purpose of this research study is to examine the correlation between Texas Regional Collaboratives (TRC) professional development for science teachers and student achievement. This project also serves a dual purpose of providing schools with a Texas Essential Knowledge and Skills (TEKS)-correlated diagnostic measure of student achievement in science which can be used to guide instruction. Nonetheless, as with any study the foundation is built upon valid and reliable scores. Therefore, we overview the methodology used to develop the grade 4 Science field test and provide evidence of score validity and reliability.

SAMPLE AND METHODS

Data were collected from 2,523 students which spanned 4 TRC Collaboratives, 43 schools, and 110 teachers. The majority of the students (51%) were White/Caucasian. Thirty-one percent were Hispanic/Latino and 12% were Black or African American. The sample had a fairly even spilt between males and females (see Table 1.). Approximately ten grade 4 Regional Collaborative teachers from each of the four Regional Collaboratives site (Education Service Center 14, North Central Texas College, Texas A&M University-Texarkana, and Our Lady of the Lake University) were targeted for participation. The control group consisted of at least ten colleagues in grade 4 that have not had training through the TRC. These colleagues were from the same school or district or a school/district with similar demographics. Participants were asked to administer the grade 4 Field Test as a posttest at the end of the school year. Teachers were notified via email to login and complete the Teacher Professional development questionnaire. Data was available to teachers on a student basis and disaggregated by TAKS objectives. Student names were masked to researchers; that is no individually identifiable student data was collected. Teacher names were also masked. Data was compiled and presented in aggregate form for the purpose of generalizing across the Collaboratives.

Table 1. Student Demographics

Gender	f	%
Male	720	51%
Female	716	49%
Student Ethnicity	f	%
American Indian/Alaska Native	2	.001%
Asian	9	.006%
Black or African American	168	12%

Table 1. Student Demographics cont.

Student Ethnicity	f	%
Hispanic/Latino	451	31%
White/Caucasian	734	51%
More than one race	72	.05%
Total	1436	100%
*Note. The student demographic information is base teacher estimates. Students were not asked to identify their race.		

The field test is based on TEKS (student expectations) and models the TAKS test by measuring four performance standards: (1) Nature of Science; (2) Life Science; (3) Physical Science; (4) Earth Science. The grade 4 Science field test has 35 items and is web based with online reports that help science educators identify strengths and weaknesses in their science program. This tool aids science educators as well as science education researchers in understanding the relationship between professional development for science teachers and student achievement.

A mixed-methods approach was used in the project. A science assessment specialist developed 40 items that were piloted with a group of students. Content and construct validity were initially addressed through a review panel of three master teachers/science education specialists from the University of Texas at Austin. The panel reviewed each item for scientific accuracy and correlation to the intended student performance expectation (TEKS). Particular emphasis was placed in ensuring that each item was a valid measure of the science content and process skills addressed in its correlated TEKS and that items measured the TEKS at the depth and complexity appropriate for the grade level addressed. As reflected in the state of Texas standardized science test in Grade 5, performance standards did not have the same number of

items (Nature of Science – 13 items, Life Sciences – 9 items, Physical Sciences – 9 items, Earth Sciences - 9 items). Note the number of items is greater for the Nature of Science objective. This is an indication of the emphasis placed on this objective. A performance standard was created by summing the respective items.

Researchers piloted 40 items with 135 students. A student focus group was conducted for the purpose of reviewing each assessment item and providing feedback regarding the completeness of the assessment as a whole, as well as clarity and reasonableness of individual items. The instrument was further refined based on item analysis. As a result, 5 items were dropped from the test due to lack of discrimination and ambiguous measures. The average item difficulty for the test was .56, which meets acceptable psychometric standards. Cronbach alpha internal consistency reliability was computed for each set of items that defined a performance standard.

Items are used to collect data from which the coherence and utility of the idea (i.e., variables) are determined. Validity deals with making and evaluating inferences drawn from test scores. No test is valid or invalid in itself. Only its use in some application merits a designation of validity (Allen & Yen, 1979; Pedhazur & Pedhazur -Schmelkin, 1991). The classic reliability model views a test score as having two additive components, the "true" score and a "random" error $X = T + E$. The error is defined as unrelated to the true score and unrelated to the error that would occur in another measurement of the same attribute. The true score is defined as "the average score taken over repeated independent testing with the same test—[it] is a theoretical idea" (Allen & Yen, 1979, p. 60).

ANALYSIS

Validity and reliability of scores are ubiquitous terms in social science measurement. According to APA standards (1985), "Validity is the most important consideration in test evaluation" (p. 9). The qualitative aspects of validity are conceptual, and the quantitative aspects are numerical. The concept of a variable is fundamental to validity. A variable is the unit of analysis of scientific inquiry and is intended to be a unidimensional manifestation of one clear idea, e.g., science ability.

Confirmatory factor analyses were utilized to test the constructs and calculate validity coefficients (Schumacker& Lomax, 2004). For a given set of response variables $x_1,...x_q$ one wants to find a set of underlying latent factors $\xi_1,...\xi_n$, fewer in number than the observed variables. These latent factors are supposed to account for the intercorrelations of the response variables in the sense that when the factors are partialed out from the observed variables, there should no longer remain any correlations. This leads to the model

$$x_i = \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + ... + \lambda_{in}\xi_n + \delta_i, \qquad (1)$$

where $\delta_{i,}$ the unique part of x_i, is assumed to be uncorrelated with $\xi_1, \xi_2, ..., \xi_n$ and with δ_j for $j \neq i$. The unique part $\delta_{i,}$ consist of two components: a specific factor s_i and a pure random measurement error e_i. The term δ_i is often called the measurement error in x_i. "In a confirmatory factor analysis, the investigator has knowledge about the factorial nature of the variables that he/she is able to specify that each measure χ_i depends only on a few of the factors ξ_j" (Joreskog & Sorbom, 1993, pp.23). The data included item level responses across four

science related areas of proficiency. The analysis was conducted using the respective set of performance indicator variables. The percent variance explained for each science standard by the respective set of performance indicators was reported along with a goodness of fit index. The goodness of fit indices (Chi –Square=4.25, df = 2, p-value=0.11966, RMSEA=0.021) indicated that the percent of variance-covariance among the performance standards is reproduced by the unidimensional confirmatory factor model.

Table 2 indicates the student science performance standards, validity and reliability score results. The science standards had the following number of performance standards:

- Standard 1 - 11 performance indicators
- Standard 2 - 8 performance indicators
- Standard 3 - 7 performance indicators
- Standard 4 - 9 performance indicators

Cronbach alpha internal consistency reliabilities ranged from .33 to .57 (overall .77). Factor loadings or validity coefficients ranged from .56 to .77. Composite factor reliabilities based on the set of performance standards ranged from .73 to .87. Percent factor (construct) variance explained ranged from .32 to .60.

Table 2. Grade 4 Science Performance Standards: Validity and Reliability of Performance Standards (n = 2,523)

Standard	Indicator (# items)	Cronbach Alpha	Validity Coefficient	Measurement Error	Percent Variance (1-error)
TAKS 1: Nature of Science	11	.57	.77	.40	.60
TAKS 2: Life Science	8	.44	.66	.56	.44
TAKS 3: Physical Science	7	.33	.56	.68	.32
TAKS 4: Earth Science	9	.43	.64	.59	.41

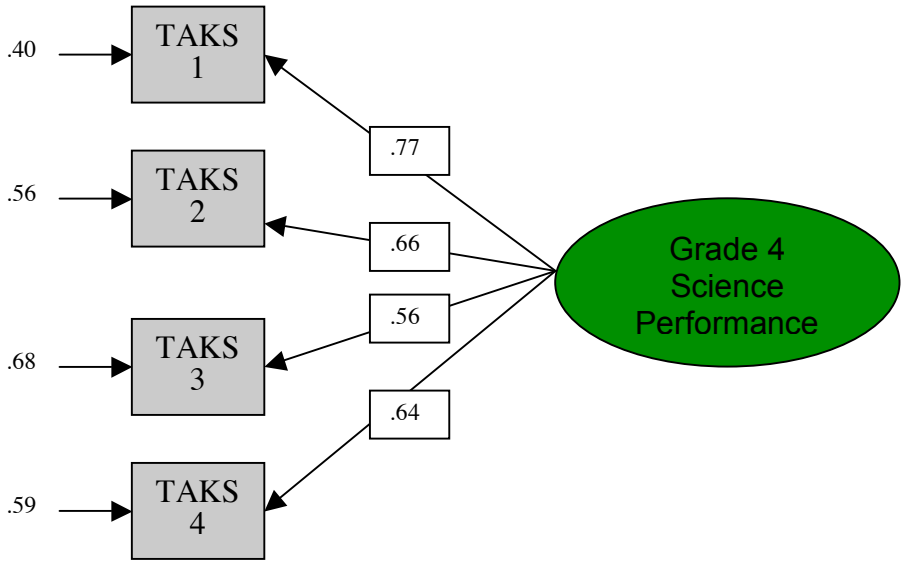

Figure 1. Confirmatory Model for Science Grade 4 (Chi –Square=4.25, df = 2, p-value=0.11966, RMSEA=0.021).

CONCLUSIONS

In summary, the grade 4 Science Field Test measures four science standards and defines 44% of the factor construct variance (i.e., Grade 4 Science Performance). The four science standards define a unidimensional construct, which is statistically supported by the confirmatory factor model fit indices. Consequently, we can interpret the validity coefficients (factor loadings) as true score weights. TAKS 1 is the most valid indicator of Grade 4 Science Performance, followed by TAKS 2, TAKS 4, and TAKS 3. All validity coefficients are of respectable magnitude and the overall test score reliability is .77. The Grade 4 Science Field Test provided a valid and reliable score assessment framework for the measurement of science performance in Grade 4. This was confirmed by the confirmatory factor model. The grade 4 Science field test provides researchers and educators a diagnostic measure of science performance. The overall measure will serve as a viable assessment of science performance in the 4th grade.

RECOMMENDATION

- Improve the score reliability of TAKS 3.
- Improve the overall score reliability.
- Pilot new test items to further the score validity and reliability.
- Conduct a quasi-experimental or experimental study to assess the science performance across an academic year.
- Investigate the correlation between professional development metrics and student performance.

REFERENCES

Allen, M. J., & Yen, W. M. (1979). Introduction to measurement theory. Monterey, CA: Brooks/Cole .

Joreskog, K. & Sorbom, D. (1993). Lisrel 8: Structural Equation Modeling with the SIMPLIS Command Language. Mahwah, NJ: Erlbaum.

Pedhazur, E. J., & Pedhazur-Schmelkin, L. (1991). Measurement, design, and analysis: An integrated approach. Hillsdale, NJ: Erlbaum.

Ramsay, M.C., & Reynolds, C. R. (2002). Development of A Scientific Test: A practical guide. Comprehensive Handbook of Psychological Assessment, Volume 1, Intellectual and Neuropsychological Assessment Gerald Goldstein (Editor), Sue R. Beers (Editor)

Schumacker, R. E., & Lomax, R.G. (2004) A beginner's guide to structural equation modeling, 2nd Edition. Mahwah, NJ: Erlbaum.

APPENDIX

Grade 4 Science Field Test- Item Map

TAK	TEK	TEK	Ans	Item	Objective
1	4.1	A01	2	01	Demonstrate safe practices during field and laboratory investigations
1	4.2	D02	2	02	Communicate valid conclusions
2	3.8	B01	4	03	Observe and identify organisms with similar needs that compete with one another for resources such as oxygen, water, food, or space
3	4.7	B02	2	04	Conduct tests, compare data, and draw conclusions about physical properties of matter including states of matter, conduction, density, and buoyancy
4	4.6	A01	3	05	Identify patterns of change such as in weather, ... and objects in the sky
4	4.11	C01	2	06	Identify the Sun as the major source of energy for the Earth and understand its role in the growth of plants, in the creation of winds and in the water cycle
1	4.2	A01	4	07	Plan and implement descriptive and simple experimental investigations including asking well-defined questions, formulating testable hypotheses,
1	4.2	C01	2	08	Analyze and interpret information to construct reasonable explanations from direct and indirect evidence
4	4.10	A01	2	09	Identify and observe effects of events that require time for changes to be noticeable including growth, erosion, dissolving, weather, and flow
1	4.2	D01	3	10	Communicate valid conclusions
2	3.8	C01	1	11	Describe environmental changes in which some organisms would thrive, become, ill or perish
1	4.2	E01	3	12	Construct simple graphs, using tools [including computers] to organize, examine and evaluate information
4	3.11	C01	3	13	Identify the planets in our solar system and their position in relation to the sun
1	4.2	E02	1	14	Construct simple tables, maps, and charts using tools [including computers] to organize, examine and evaluate information

TAK	TEK	TEK	Ans	Item	Objective
1	4.3	A01	1	15	Analyze, review [critique] scientific explanations, including hypotheses and theories, as to their strengths and weaknesses using scientific evidence and information
4	3.6	B01	3	16	Identify that the surface of the Earth can be changed by forces such as earthquakes and glaciers
1	4.4	A01	2	17	Collect and analyze information using tools including calculators, microscopes, [cameras, sound recorders, computers,]hand lenses, rulers, thermometers, compasses, balances, [hot plates] meter sticks, timing devices, magnets, collecting nets, and safety goggles.
2	3.8	A01	4	18	Observe and describe the habitats of organisms within an ecosystem (2.9B)
4	3.11	A01	1	19	Identify and describe the importance of earth materials including rocks, soil, water, gases of the atmosphere in the local area and classify them as renewable, non renewable or inexhaustible resources.
2	3.8	D01	3	20	Describe how living organisms modify their physical environment to meet their needs such as beavers building a dam or humans building a home
2	4.9	A01	4	21	Distinguish between inherited traits and learned characteristics Identify and provide examples of inherited traits and learned characteristics
1	4.3	B01	2	22	Draw inferences based on information [related to promotional materials] for products and services
3	3.6	A01	2	23	Measure and record changes in the position and direction of the motion of an object to which a force such as a push or pull has been applied.
2	4.8	B01	1	24	Compare adaptive characteristics of various species
2	4.8	A01	3	25	Identify characteristics that allow members within a species to survive and reproduce (2.9A)
3	4.6	A01	3	26	Identify patterns of change such as in weather, metamorphosis, and object in the sky.

TAK	TEK	TEK	Ans	Item	Objective
3	4.7	A01	1	27	Observe and record changes in the states of matter caused by the addition or reduction of heat
3	4.6	B01	3	28	Illustrate that certain characteristics of an object can remain constant even when the object is rotated (symmetry reflections)
4	4.11	B01	4	29	Summarize the effects of the oceans on land
3	4.7	B01	1	30	Conduct tests, compare data, and draw conclusions about physical properties of matter including states of matter, conduction, density, and buoyancy
1	4.2	A02	2	31	and selecting and using equipment and science
4	4.10	B01	1	32	Draw conclusions about “what happened before” using fossils or charts and tables.
3	4.7	A02	2	33	Observe and record changes in the states of matter caused by the addition or reduction of heat
4	4.11	A01	1	34	Test properties of soils including texture, capacity to retain water, and ability to support life
2	4.6	A01	1	35	Identify patterns of change such as in weather, metamorphosis, and objects in the sky

http://www.i2-k.com/GR4SCIENCE/ItemMap.htm

Teacher Online Survey Instructions:

1. Please complete the survey a few days prior to taking students to the computer lab for their online test.
2. Go to http://www.i2-k.com/cgi-bin/index.cgi
3. Select "Teacher".
4. Enter your username and password.
5. Take the questionnaire. NOTE: If you do not complete every question in the survey you cannot receive student id numbers. Every textbox must contain an answer even if it zero "0".
6. Click on "Create Student Accounts".
7. Select "View Student Accounts".
8. Print a copy of this page by going to file/print.
9. Write student names next to the ID numbers. IMPORTANT NOTE: Students will need this number when they take their online test. Safeguard this list, as it will be the only means to associate test results to individual students. Test results will be reported back to you with an ID number only.
10. Select "View Student Accounts" and complete demographic data for each student by selecting “Set” located in the far right column of the table. This data is required.

After students have completed the test.

11. View Online Reports (student, class average, campus average, overall test average) after students have completed the test. The summary may indicate areas of weaknesses or strengths in your science program.

Grade 4 Science Field Test - Instructions for Students:

To the teacher:

Read the instructions to the students allowing them time to complete each step. Your students will need to be familiar with drop down menus to complete the student information. Print out a copy of the student instructions and tape it next to the computer.

1. Go to http://www.i2-k.com/cgi-bin/index.cgi
2. Select "student".
3. Enter your student id number (This is the number your teacher assigned to you).
4. Complete the rest of the information and the test.
5. Do NOT use the back arrow button on your browser.
6. If you want to go back to a question, write down the question number on a piece of paper. You may review (Go Back to) questions at the end of the test by clicking on the review (Go Back) button.
7. Once you have checked your answers. At the bottom of the page "Click here to complete and save your test". Note test score and exit browser.

We will use the "allowable accommodations" for TAKS as a general guide 2006 Coordinator Manual, Texas Student Assessment Program, p. 32-38.

http://www.tea.state.tx.us/student.assessment/resources/guides/coormanual/index.html

If you have any questions about this survey, please contact me at Sherron@i2-k.com or call (512)-921-1630. Your assistance is greatly appreciated, and we thank you in advance for your thoughtful answers and prompt reply.