Brief Report to the Teachers Academy for Mathematics and Science: Psychometric Properties of A New Test to Assess Basic Skills in Mathematics and Science of Teachers and an Update Analysis on the Teacher Attitude Survey: School Year 2002-2003

Prepared by

Kathryn E. H. Race
Race & Associates, Ltd.
4430 N. Winchester Avenue
Chicago, IL 60640
race_associates@msn.com

Prepared for

Teachers Academy for Mathematics and Science
501 W. 35th Street
Chicago, IL 60616
www.tams.org

March 2003

Brief Report to the Teachers Academy for Mathematics and Science: Psychometric Properties of A New Test to Assess Basic Skills in Mathematics and Science of Teachers and an Update Analysis on the Teacher Attitude Survey: School Year 2002-2003

Kathryn E. H. Race
Race & Associates, Ltd.

The purpose of this short report is to provide a brief summary of the results of analyses on the psychometric properties (i.e., internal reliability and construct validity) of two measurement tools. These tools are used by the Teachers Academy for Mathematics and Science to support outcomes evaluation efforts of their intensive professional development program for elementary teachers, in particular to assess intermediate outcomes related to teacher knowledge, and attitudes, and in support of measures of instructional practices by participating teachers.

These measures are:

1. A New Teacher Basic Skills Test
2. Teacher Attitude Survey

The basic skills test in mathematics and science has been redesigned based on the work from two external consultants, designated program staff and internal evaluation staff. The teacher attitude survey was modified based on the results of psychometric analyses conducted on data from the 1999-2000 and 2001-2002 school years. In particular, a new scale designed to assess teachers as learners was added. Additional items or revised items were included to support existing scales or to help clarify previously ambiguously worded items (Race, 2000; 2001).

Background

The Teachers Academy for Mathematics and Science is a non-profit organization located in Chicago. The Academy "is an autonomous alliance of leaders from education, government, science, mathematics, business, and the community" (Teachers Academy for Mathematics and Science, 1998, p.4). Since late in 1990, the Academy has offered an intensive 3-year professional development program in mathematics and science to meet the needs of under-prepared elementary school teachers in Chicago and other school districts within the state of Illinois.

Each teacher who participated in the Academy's intensive professional development program beginning in school year 2002-2003 from participating schools from the Chicago Public School (CPS) was asked to complete four measurement tools. These are

a basic skills test in mathematics and science, a teacher attitude survey, a self-report assessment regarding their technology skills, and a brief biographical form. The psychometric properties of only the basic skills test and the attitude survey were analyzed.

As has been done in previous analyses, the internal reliability of the overall basic skills test was assessed using Cronbach's alpha (Cronbach, 1951). To assess the construct validity of each instrument, a factor analysis was conducted based on principal components analysis and varimax rotation (Kleinbaum, Kupper & Muller, 1988).

To support the solution obtained for the basic skills test, principal axis factoring (another method to extract factors) was also used to test the robustness of the obtained solution in order to seek factor groupings that remained consistent across models. To help guide decisions on the number of factors to extract, a parallel analysis for each instrument tool, based on a matrix of randomly generated numbers with similar parameters regarding sample size, number of items, and response category options, was conducted as well (Thompson & Daniel, 1996).

Basic Skills Test in Mathematics and Science (See Appendix A.)

A previous version of a content knowledge test in mathematics and science was shown to very poorly represent the content presented in the intensive professional development program for elementary school teachers offered by the Teachers Academy (Whitkanack, 2001). As a result, a task force comprised on two external consultants, one in mathematics and one in science, plus program staff and the director of evaluation work to create a test that was more sensitive to program-related content and covered "big ideas" in mathematics and science addressed in the program. This work culminated in a revised basic skills test described as follows. As revised, the test contained 35 items, some with component parts others self-contained. These items were comprised of TIMS items (that is, Teaching Integrated Mathematics and Science), new mathematics items (new in the sense of not being included in previous test versions) and new science items (also new in the same sense of not being included in previous test versions). Based on a 1-point per part scoring scheme, a total of 71 points were possible. An example of the test is not provided, since it is currently used to grade progress in mathematics and science by participating teachers.

Item Analysis

The first step in analysis was to compare each item to the total-test score as well as to compare inter-item correlations for all 71 item-parts. Results of this analysis suggested the elimination of 5 items (a total of 8 points) largely due to the fact that these items tended to reflect a negative relationship. That is, these items tended to be answered correctly by participates who did poorly on the test; and conversely, those who did well on the test tended not to answer these correctly. This, of course, is counter to what is intended. Thus, these items were deleted.

With these item-parts deleted, a total of 63 points was possible. Based on this total, the average overall test score was 31.28 (standard deviation = $\pm$ 12.79). The internal reliability of the overall test was high, alpha =.93, suggesting statistical support for using this total score as a dependent measure in future data analyses.

Factor Analysis

A factor analysis, using principal components and varimax rotation, was used to explore whether the test measures more than one construct. Visual inspection of a scree test suggested that a 3 to 5 factor solution was possible, since there was a break after factor 3, and again after factor 5. Results of a parallel analysis suggested that a solution containing more than five factors were possible as well (up to 10). On the basis of a conservative approach, solutions retaining 3, 4, and 5 factors were explored. These analyses suggested that factor 5 was comprised of only a few items and probably did not represent a stable nor well-defined construct. Exploring a 4-factor solution suggested sufficient overlap in factors 3 and 4 to legitimately question whether these represented separate constructs.

A 3-factor solution seemed very appealing since the test as comprised, contained TIMS-items, new mathematics items, and new science items. Inspection of this solution, however, did not suggest that these three factors were configured in clear and unambiguous constructs. More specifically, the 3-factor solution showed evidence that the TIMS items were split across two factors. This was also the case for the new mathematics items, that is, the new mathematics items were split across the first two factors. This was not the case for the third factor, however, which contained all of the new science items.

On this basis, a two-factor solution was accepted, with eigenvalues, of 12.97 and 3.84, respectively. Together these two factors accounted for approximately 27% of the common variance. These findings were supported by a principal axis factoring; thus, suggesting the robustness of this overall model. The individual items, which comprised each of these factors, are shown in Appendix A.

Internal Reliability

Factor 1, labeled TIMS and New Math Items, was comprised of 44 item-parts. The average score based on this sub-total was a mean of 19.77 (standard deviation = 9.80). Factor 2, labeled New Science Items, was comprised of 19 item-parts. The average score for this sub-total was 11.51 (standard deviation of 4.71).

Each of these sub-totals has good internal reliability, for Factor 1 an alpha = .93, and for Factor 2 an alpha = .86. This provides statistical support for the use of these sub-totals, which could provide separate progress by teachers on TIMS and mathematics, and again for science.

Teacher Attitude Survey (See Appendix B.)

The modified attitude survey consisted of a total of 67 items. These items were divided into the following sections: general statements about teaching (13), teaching and learning (11), instructional strategies (14), understanding and confidence in teaching mathematics and science (9), methods and approaches to teaching (9), and using computers and other technology in the classroom (11). The response categories used to rate each item were based on a 5-point, Likert-type scale (Likert, 1932) from 1 = strongly disagree to 5 = strongly agree (an option of 3 = neutral was added). A total of 381 teachers returned a survey of which 300 or 79% were used in this analysis.

Preliminary item and factor analyses of this modified scale resulted in the elimination of 16 items. This resulted in 51 retained items. Overall reliability estimates were not conducted on the overall instrument because of clear evidence that the instrument is comprised of multiple dimensions.

Construct Validity. The decision to retain five factors from the modified survey was supported by the results from previous analysis, plus the addition of a new scale. A scree plot of the eigenvalues associated with identified factors suggested a break after 5 factors. The results from a parallel analysis suggested the retention of five factors as well. Using principal components and varimax rotation, a five-factor solution was accepted which accounted for 55% of the common variance, with eigenvalues of 14.58, 4.73, 3.99, 2.80, and 1.87, respectively. The results of this analysis strongly suggests the robust nature of the four original scales obtained from previous analyses (Race, 2001).

The new scale, *Teachers as Learners*, however, shows considerable overlap and non-independence with one of the original scales, that is, *Inquiry-based Instructional Strategies*. Elimination of dual cross-loadings by items on these two factors did not occur until the solution was rotated using promax, an oblique rotation. The results of these analyses are shown in Tables B-1 and B-2 in Appendix B for the oblique rotation using promax and the rotation based on a varimax approach, respectively.

The results of these analyses suggest that the instrument can be used to measure five factors as shown in Table 1 with the first two factors overlapping considerably, reflecting non-independent, and shared variance. The results of the varimax rotation, which attempts to maintain an orthogonal, independent solution (see Table B-2 in Appendix B), shows the degree of overlap between the *Inquiry-based Instructional Strategies* and *Teachers as Learners* scales. In other words, these two scales are highly connected and do not measure unique aspects of attitudes toward teaching practices and pedagogy. Moreover, the shared variance between these two scales has serious limitations for subsequent analyses, in particular multiple regression analyses

Despite the emergence of this new scale, its lack of independence with the other scales suggests a redundancy in the attitude scales that is really not necessary or helpful. A more parsimonious solution would result in the retention of only the four original scales. That is, for future use of this attitude scale with *new* cohorts of teachers, I would

Table 1
Summary of Results from the Assessment of the
Teacher Attitude Survey Samples from Two Different School Years

Teacher Attitude Survey	Chicago Public Schools 2002-2003	East St. Louis – Joliet 2000-2001
Number of surveys used in analysis	300	320
Total number of items	67 items (16 items eliminated)	56 items (6 items eliminated0
Summary of Factor Analysis		
Factor 1. Inquiry-based Instructional Strategies	14 items; alpha = .94	23 items; alpha = .91
Factor 2. Teachers as Learners	9 items; alpha = .91	-----
Factor 3. Reluctance to Use Traditional Methods and Approaches to Teaching	9 items; alpha = .87	8 items; alpha = .83
Factor 4. Understanding and Confidence in Teaching Mathematics and Science	8 items; alpha =.84	7 items; alpha =.80
Factor 5. Using Computers and Technology in the Classroom	11 items; alpha = .83	11 items; alpha = .81

Note. The inclusion of the new scale, *Teachers as Learners*, seriously limits the independence and unique aspects of the constructs measured by the original four scales

recommend that the four original scales be retained and the first 24 items of this modified instrument be eliminated. This would include the elimination of this new scale as well. This would reduce the total scale to 43 items; the reduced length having the added advantage of a shorter, easier to complete instrument by participating teachers. Using only four scales would also enhance comparisons between this new cohort of teachers and previously participating teachers from East St. Louis and Joliet.

It should be noted that only 79% of the teachers completed this present survey versus an 88% completion rate for a past survey of comparable length (Race, 2000; 2001). It is possible that this difference in completion rate may be due to the increased length of the instrument as well as methodological differences in gathering this information from previous cohorts of teachers.

Reliability. The alpha coefficients of the original scales and the new scale suggest good internal reliability for each identified scale, .94, (.91), .87, .84, and .83, respectively (see Table 2).

Summary

The results of these analyses suggest that both the basic skills test and teacher attitude survey have good psychometrics relative to internal reliability and construct validity.

The basic skills test may be used, when the recommended items are eliminated, based on using a total score of a simple sum of all retained items. Or, two sub-scores can be used based on the simple sum of all items in Factor 1 (TIMS and *new* math items) and a simple sum of all items in Factor 2 (*new* science items). But, all three of these measures should not be used as dependent measures in the same analysis because they are not independent of each other. Using Factor 2 sub-scores may be particularly instructive in gauging teacher progress in science during the course of the intensive professional development program.

Although, it might also be useful to have separate sub-scores for TIMS items, math items, and science items such a practice is not recommended since the factor structure of the overall test does not support this division. If this is desirable, then it is recommended that the test be restructured to meet this intent. As it stands currently, however, the two sub-scores may be very useful, in and of themselves, and may be a practical metric to gauge teachers progress through the intensive professional development program without further test development.

The results of analysis on the teacher attitude survey are a mixed positive. On the one hand, the original four scales were robustly confirmed by the present analysis. The new scale, *Teachers as Learners,* however, is highly ambiguous as a construct at this stage and it is not recommended that it be used in its present form. Since the original scales have been shown to be sensitive to before-after attitude changes, it would seem that these four scales are sufficient.

Finally, the methodology used to obtained the basic skills test and attitude survey data needs to be reviewed. That is, there is a sizeable discrepancy between the number of completed basic skills test (a total of 500) and the number of teacher attitude surveys that were received (a total of 381). More specifically, the basic skills test was administered during a *pre*-program session, whereas the teachers were allowed to take their attitude survey home and return with completed surveys at the start of the first program session. For this or other reasons, 24% of the teachers who completed a basic skills test did not complete the teacher attitude survey. Unless, this represents an actual drop in attendance from *pre*-program to program session attendance, this methodology should probably not be repeated for future cohorts of teachers.

References

Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16,* 297-334.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariable methods.* Belmont, CA: Duxbury.

Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 140,* 1-55.

Race, K. E. H. (2000, February). *Preliminary assessment of psychometric properties for four measurement tools used in the evaluation efforts of the Academy program: School year 1999-2000.* Chicago: Teachers Academy for Mathematics and Science.

Race, K. E., H. (2001, February). *Update on psychometric properties for four measurement tools used in evaluation efforts of the Academy program: School year 2000-2001.* Chicago: Teachers Academy for Mathematics and Science.

Teachers Academy for Mathematics and Science. (1998, June). Strategic plan 1998-2003. Chicago: Author.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*(2), 197-208.

Whitkanack, D. (2001, September). *The development of a matrix to align the basic skills test in mathematics and science and professional development sessions.* Winfield, IL: Author.