The Internal Consistency of the *Mankato Scale*: Laying the Foundation for Valid Professional Development Decisions

Antonia A. D'Onofrio
Center for Education
Widener University
United States of America
Antonia.A.D'Onofrio@widener.edu

Kathleen A. Bowes
Center for Education
Widener University
United States of America
Kathleen.A.Bowes@widener.edu

Abstract. The *Mankato Scale* has been used since 1992 for staff development purposes. By describing teacher strengths and deficiencies, it points the way for professional development programs to encourage greater infusion of technology into classroom instruction. The absence of empirical information about the reliability of the *Mankato Scale* is conspicuous. Yet professional development programs have an ever-present need to establish baseline conditions of ability and skill, and to create professional development plans that can be monitored. This investigation demonstrated that the *Mankato Scale* is an internally consistent and reliable tool. Hence, the scale can be legitimately used as part of professional development for grouping participants by level of competence to obtain maximum results from training. Having such a reliable tool can also assist administrators to monitor the progress of their training initiatives.

Introduction

The *Mankato Scale* has been used since 1992 for staff development purposes (Johnson, 1999). By describing teacher strengths and deficiencies in technology integration, it points the way for professional development programs to encourage greater infusion of technology into classroom instruction (Johnson, 1997). The absence of information about the internal reliability and statistical validity of the *Mankato Scale* is conspicuous (Johnson, personal communication, December 3, 2002). Credible evaluations of professional development programs are hampered when the assessment tools used to measure the before and after abilities of participants are not internally reliable (Cook & Campbell, 1979; Wilson, 1992-1994). The aim of this investigation was to discover whether the *Mankato Scale* is internally reliable, and whether the scale meets objective criteria required to demonstrate minimal levels of construct validity. This is a necessary, preliminary step toward establishing whether the *Mankato Scale* can be used to make legitimate professional development decisions based on its assessment of teachers' integration of technology.

Rationale

The validity of any assessment tool is initially dependent upon the ability of scale items to measure underlying knowledge, skills, or dispositions consistently (Goldstein & Hersen, 1984; Jaeger, 1990). Therefore, one cannot say that the *Mankato Scale* assesses teachers' levels of technology integration reliably without knowing that the scale measures their use of technology consistently. Those responsible for the professional development of teachers, and the teachers themselves, may discover that information gleaned from the *Mankato Scale* motivates them to continue to learn and grow professionally (Kinsey & McKenna, 1999; McKenzie, 1994; Wepner & Tao, 2002). However, absent any evidence that the *Mankato Scale* provides information that is internally consistent (i.e., internally reliable or stable), the scale's ability to contribute to the continuing growth and development of teachers is limited.

Why is it important to understand the reliability and validity of any scale? Generalizability, or the ability to draw conclusions based on information gained from a scale, is essential to establishing the value of a scale for an array of uses, including simple descriptions of study participants based on assessment results. Generalizability depends on two related scale characteristics: the internal consistency of scale items and the construct validity of the scale. Internal consistency, an important index of a scale's reliability, determines how well items that comprise a scale contribute to internally stable assessment (Thompson & Daniel, 1996; Trochim, 2002). Construct validity determines whether a scale measures the underlying construct (e.g., level of technology integration), that it claims to measure. Thus, empirical demonstrations of internal consistency are the first step in being able to claim that the *Mankato Scale* measures levels of technology integration in a meaningful way (Kerlinger, 1979; Nunnally, 1978; Trochim, 2002).

Internal Consistency

Researchers need to be able to describe the extent to which items are empirically related to one another, and thus comprise a scale. Decision makers need to know which scale items assess technology integration precisely. When these two characteristics are present, then a collection of items can be considered a scale whose overall score makes objective sense. Establishing the internal consistency among items on a scale can be accomplished by calculating item to item correlations, item to total score correlations (split-half reliability or Cronbach's alpha), scalability estimation (Guttman's Q coefficient of scalability or Rasch modeling), and factor analysis. All these techniques are able to demonstrate that a scale is an operational construct (Guttman, 1950; Linacre, 2000; Palmquist, 2005).

Only a few studies have investigated the internal consistency of various tools and frameworks for assessing technology integration. Moersch (22004) reports that 50 items comprising the *Levels of Technology Implementation Questionnaire* (LoTiQ) define 3 subscales (Levels of Technology Use, Personal Computer Use and Stages of Instructional Practice) having internal consistency estimates of .75, .81 and .74, respectively. The *Technology Ability Perception Self-report Instrument* explores 6 subscales, as follows: Technology Operations and Concepts, Planning and Designing Learning Environments, Teaching, Learning and the Curriculum, Assessment and Evaluation, Productivity and Professional Practice, and Social, Ethical, Legal and Human Issues. Cronbach's alpha coefficients computed for these scales are .91, .86, .85, .85, .79, and .85, respectively (Pawloski, 2003). Other well known measures of technology integration (including the *High Plains Regional Technology in Education Consortium* (HPR*TEC*) Profiler*, the *California Technology Assistance Project (*iAssessment*)*, *enGauge*, *Taking A Good Look at Instructional Technology* (TAGLIT*)*, *Southeast Initiatives Regional Technology in Education Consortium* (SEIR*TEC), and *Learning with Technology Profile Tool* (NCREL) have not provided psychometric information about their assessment systems.

Two well-established scales have explored teachers' attitudes toward technology and publish related reliability statistics on the web. Knezek and Christensen (1998) investigated the internal consistency of *The Teachers' Attitude Toward Computers* (TAC) for various item-scale configurations, identified through factor analysis. Sixteen factors or scales yielded Cronbach's alpha values ranging from .74 to .98; ten factors yielded alpha coefficients ranging from .81 to .97; nine factors yielded alpha coefficients ranging from .82 to .94; and 7 factors yielded alpha coefficients ranging from .90 to .98. Knezek and Christensen have in addition reported values of Cronbach's alpha ranging from .91 to .96 for 10 scales comprising the *Teachers Attitudes Toward Information Technology* (TAT). This is a scale that explores attitudes toward email, the world wide web, multimedia, and productivity tools.

The Relationship between Internal Consistency and Construct Validity

If one can establish the internal consistency among items on a scale, then one can say that the scale is a scale because its items correlate with each other and with an overall score (Fink, 1995; Kerlinger, 1979). Construct validity provides meaningful information about an underlying construct assuming that the construct has been measured consistently (Russ-Eft, 1980). Construct validity in part depends on a scale's content validity or whether item content is logically related to the underlying concept measured by a scale -- e.g., items appear to measure technology integration, and item content can be indexed to idealized outcomes of technology integration (Trochim, 2002). Construct validity also requires statistical correspondences between items (convergent validity), and, conversely, statistical divergence (discriminant validity) which shows that items are not measuring different underlying content, skills or attitudes (Gorsuch, 1983; Nunnally, 1978). Factor analysis is commonly used to demonstrate statistical convergence and statistical divergence of scale items (Cattell, 1978; Gorsuch, 1983; Kerlinger, 1979; Nunnally, 1978).

According to Moersch (2000) "factor analysis revealed LoTiQ levels to be significantly correlated to Personal Computer Use (PCU) (r = .579). In addition, LoTiQ levels were found to be significantly correlated to Stages of Instructional Practice (CIP) (r =.422)." It is unclear from this description that LoTiQ levels were correlated with factor scores drawn from an analysis of scaled items, or whether LoTi, PCU and CIP were identified as statistically inferred scales through a factor analytic procedure. *The Teachers' Attitudes Toward Computers* (TAC) provides information for versions TAC 2.22 (199 items having a 16 factor structure; for TAC 3.0 (198 items having a 16 factor structure); TAC 3.2a (105 items having a 7 factor structure); TAC 3.2b (109 items and a 7 factor structure) and TAC 4.0 (151 items having an 11 factor structure). Continued revisions of the TAC indicate that the authors have experimented with item and scale content with the aim of simplifying their measurement of teacher attitudes.

Methods

Participants

One hundred and sixty graduate students attending a small multi-purpose university in southeastern Pennsylvania participated in this study. Certified classroom teachers comprised the largest number of participants; however, a smaller number included those seeking teacher certification, and others who were seeking specialty certificates. Table 1 describes participants' gender, age and teaching experience.

Table 1: Statistical Description of Study Participants (N=160)

Characteristics	Frequency	Percentage of Total
Gender		
Male	39	24.00
Female	111	69.00
Age		
Under 30 years	63	40.00
31-40 years	34	22.00
Above 40 years	58	37.00
Years Teaching		
0-1 years	59	39.00
2-5 years	48	31.80
6 or more years	44	39.00

Procedures

The study participants completed the *Mankato Scale* during class time. The permission of course instructor was granted prior to asking students to complete the surveys; and informed consent was required from the students.

Instrument

The *Mankato Scale* is self-administered and probes 16 areas of technology integration, thereby providing a profile of how extensively teachers apply technology to classroom instruction. The 16 areas (or items) correspond to

16 uses of technology: Basic Computer Operations; File Management; Network; Word-processing; Spreadsheets; Database; Graphics; Internet; Web Page Design; Telecommunications; Email; Information Searching; Ethical Use/Understanding; Presentation Skills; Instructional Software; and Technology Integration. The scale is configured so that items placed at the beginning of the assessment scale are based on more established competencies (e.g., Word-Processing). Next items probe applications of intermediate complexity (e.g., Graphics). The scale culminates with more complex problems of technology use (e.g., Responsible Use/Ethics). Teachers respond to each item by indicating the level at which they use the competency represented by each item, with possible responses ranging from Level 1 to Level 4. These four levels correspond to hypothesized levels of technology integration: (i.e., Pre-Awareness, Awareness, Mastery, and Advanced). At the lowest level or Level 1 (Pre-Awareness) a teacher indicates s/he does not use a competency. Levels 2 and 3 (Awareness and Mastery, respectively) represent intermediate levels of integration; whereas Level 4 (Advanced) represents the highest level of technology integration for any given item. Each item is scored 1 through 4 (reflecting the level selected by a teacher), yielding total scores which can vary from 16 through 64. As a result the total score achieved by teachers on the *Mankato Scale* provides two insights into the extent of technology integration: (1) overall use of technology; and (2) the extent to which a teacher integrates specific technologies included on the *Mankato Scale* with instruction.

Analyses

Cronbach's alpha (Cronbach, 1989), a measure of internal consistency based on the average of all split-half reliability estimates, was calculated in order to correlate teacher responses to individual items with their overall Mankato score. Because Johnson (1999) posits that the scale describes levels of technology integration, it was important to confirm the consistency of each of the four levels as unified dimensions. Therefore Guttman analyses (Guttman, 1950) were conducted to explore internal scale consistency for each respective level of integration. A scalability coefficient was used to determine the extent to which individuals' responses were stable statistically for each of the levels. For example, individuals who are likely to give Pre-Awareness responses to one item would consistently supply Pre-Awareness responses for other items on a scale, and *vice versa*. A scalability coefficient is used to judge unidimensionality, or consistency of responses for each level of integration-- with higher values of the scalability coefficient describing greater consistency among response patterns. Finally a principal components analysis was conducted in order to determine whether *Mankato Scale* items converge as a unified scale (*Factor Analysis Using SAS PROC FACTOR*, 1995). This technique evaluates the extent to which some items form clusters or scales of statistically correlated information, and are uncorrelated and distinguishable from other clusters of inter-correlated items.

Findings and Conclusions

Findings

The calculated value of Cronbach's alpha was .94 ($p < 01$), demonstrating a high level of internal consistency among all 16 items of the *Mankato Scale*. Values greater than 85 describe acceptable internal consistency among scale items. Guttman analyses computed the following values for the scalability coefficient *Q* for all split-halves for each of the 4 levels of technology integration. Split-half reliability for Pre-Awareness level ranged from .65 to .82 ($Q=133.86$; $df=1$; $p < .0001$). Split-half reliability coefficients for the Awareness level ranged from .52 to .65 ($Q=41.92$; $df=1$; $p < .0001$). Split-half reliability coefficients for the Mastery level ranged from .35 to .65 ($Q=13.11$; $df=1$; $p < .0003$). Split-half reliability coefficients for the Advanced level ranged from .87 to .91 ($Q=111.67$; $df =1$; $p < .0001$). These findings suggest that reported levels of integration were most consistent for participants who were more likely to supply responses associated with either Pre-Awareness or Advanced uses of classroom technologies. Response patterns associated with the level of Awareness and Mastery were statistically but relatively unstable.

A principal components analysis was computed in order to investigate the scale structure of scale items and discover whether they converged as a single scale. Table 2 indicates that this did not occur.

Mankato Scale items converged as 2 correlated dimensions of technology integration. The component loadings (values describing how well items predict components), reported on Table 2, were used to identify items which were the best predictors of the respective components, based on loading values greater than .40. Component I was predicted by 12 items whose loadings ranged from .46 to .83. Basic Computer Operations, File Management, Network Use, and Word-processing had the strongest predictive association with Component I, and this finding

Table 2: Principal Component Analysis and Alpha Factoring of Mankato Scale Items (Following a Kaiser- Varimax Rotation)

	Principal	Components		Alpha	Factors	
Mankato item	I	II	*h* ^{2}	Fundamental Applications	Advanced Applications	*h* ^{2}
	loadings	*loadings*		*coefficients*	*coefficients*	
Basic Computer Operations	*83*		*76*	*81*		*73*
File Management	*81*		*72*	*73*		*68*
Network Use	*81*		*66*	*74*		*57*
Word-processing	*71*		*64*	*68*		*61*
Spreadsheet Use	*68*		*60*	*65*		*57*
Data Base Use	*67*		*54*	*62*		*49*
Graphics Use	*64*	*42*	*60*	*69*	*44*	*57*
Internet Use	*71*	*40*	*68*		*42*	*65*
Web Page Design use		*62*	*48*	*63*	*53*	*40*
Telecommunications Use (Email)	*65*	*47*	*62*	*64*	*44*	*58*
Information Searching	*66*	*51*	*66*	*47*	*48*	*64*
Ethical Use/Understanding	*50*	*64*	*52*	*49*	*49*	*58*
Presentation Skills	*46*	*79*	*63*	*47*	*61*	*59*
Instructional Software Use		*83*	*69*		*73*	*60*
Technology Integration		*78*	*77*		*82*	*74*
Student Assessment			*66*		*70*	*56*
Eigenvalues	8.49	.98		5.50	3.97	
% Variance	53.05	6.16		34.40	24.81	
Cumulative %	53.05	59.20		34.39	59.20	

suggested that the first component primarily captured levels of teacher integration identified by the most fundamental classroom applications. Component II was predicted by 9 scale items that yielded loading values ranging from .40 to .83. Presentation Skills, Instructional Software Use and Technology Integration were the items most strongly associated with Component II. Student Assessment did not predict either component. A scree test of eigenvalues (i.e., a statistic which measures the total amount of item variance contained in each component) also established that the *Mankato Scale* items converged as two components (although the second component only approached the standard cut-off for scree tests of 1.0). Since scale items did not converge as a single scale, it was important to confirm the two scale solution suggested by the resulting 2 principal components. Alpha factoring (see Table 2) also demonstrated that *Mankato Scale* items would converge as two scales of technology integration. Once again 12 item the first factor. Computer Operations, File Management and Network Use were the best predictors of this initial factor. The first factor was entitled Fundamental Applications. Instructional Software Use, Presentation Skills, and Student Assessment were the best item predictors of the second factor, which was therefore entitled Special Applications. A scree test of eigenvalues for the 2 alpha factors confirmed that the items indeed converged as two scales or dimensions of technology integration.

Conclusions

These early analyses of the internal consistency of the *Mankato Scale* have been able to demonstrate measurement characteristics that are positive, and that point the way for continued investigation of this assessment tool. The internal consistency of the scale was confirmed by the high item to-scale alpha coefficient. Guttman analyses indicated that participants who reported their uses of technology primarily at Pre-awareness and Advanced levels supplied response patterns that were scalable (statistically stable). Response patterns at levels of Awareness and Mastery were not internally consistent (most scalabililty coefficients fell below acceptable levels of .70.). It is possible that the levels of Awareness and Mastery describe responses of individuals who are in transition as technology users and therefore working at more than one level of integration depending on the kind of technology they are learning to use. Principal components analysis and alpha factoring both indicated that *Mankato Scale* did not describe a single dimension, but rather two distinguishable dimensions of technology integration. Collectively the scale items described a second dimension of integration and application that constitutes an essential part of classroom practice. Since two alpha factors could be interpreted, the *Mankato Scale* may actually produce two scores that are meaningful for planning staff development programs.

Future studies of the measurement characteristics of the *Mankato Scale* are also needed. Replications of this study, those that explore the response patterns of teachers at a variety of levels of training and experience, would provide much needed insight into readiness of pre-service teachers, novice and veteran teachers, and teachers in specialized disciplines. Studies of longitudinal progress of teachers who acquire the skills of technology integration over long-term staff development efforts are valuable. However statistical information about the long-term stability (test-retest reliability) of the *Mankato Scale* is requisite to understanding teacher progress, and such studies have not been conducted. The *Mankato Scale* is based on self-report (teachers describe the extent to which they integration technology and instruction). Deeper understanding of technology integration would result from studies that compare teacher responses to the *Mankato Scale* with scores gathered from observational tools (e.g., LoTiQ)

Finally, this investigation took first steps toward understanding the construct validity of the *Mankato Scale*. Establishing the predictive validity of the *Mankato Scale* will depend on future research that correlates performance on this scale with assessment tools that also examine technology integration. Because of *No Child Left Behind* teachers are required to be accountable for the outcomes of classroom instruction. Technology integration is part of this process. School districts need to validate student and teacher progress on a continuum of measurable achievement. Professional development programs have an ever-present need to use assessment, in order to establish baseline conditions of ability and skill, as well as to create professional development plans that can be monitored (McKenzie, 1998). Valid assessment tools are needed to ground student, teacher and district accountability in defensible, empirical and measurable information. However, valid assessment demands assessment tools that are reliable (stable and consistent); and statistical validity as well as content validity must be evident.

At this time, too few surveys, tests and frameworks for judging levels of technology integration can withstand this kind of scrutiny. Many commonly used assessment tools are not publicly accessible for preliminary statistical evaluations in established journals, psychometric resources or the world wide web. Many assessment tools, which are available through proprietary web sties, require membership identities in order to view test content. In some cases, items are customized for specialized assessments and are drawn from larger banks of items known only to the publishers. Often it is not clear whether studies of reliability and validity have even been conducted. Clearly assessments of technology integration are expensive to develop and produce, and authors and publishers want to protect their intellectual property. Nonetheless, when the value of intellectual property outweighs access to information about reliability and validity (not to mention the content specifications used to develop an assessment tool), then the interests of scientific inquiry are not well served. The legitimacy of professional staff development decisions, whether educators realize it or not, is directly related to the empirical soundness of its assessment tools and a transparent context of inquiry that allows these tools to be evaluated.

References

CTAP Online iAssessment & Technology Use Survey. Retrieved January 5, 2005 from http://www.scusd.edu/education_technology/ctap_online_iassessment.htm

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Plenum.

Cook, T. D. & Campbell, P.T. (1979). *Quasi-experimentation: Design and analysis, issues for field settings.* Boston, MA: Houghton-Mifflin.

Cronbach, L. (1989). Construct validation after thirty years. In R. L. Linn (ed.). Intelligence*: Measurement, Theory and Public Policy* (pp. 147-171). Urbana: University of Illinois Press.

enGauge. Retrieved January 5, 2004 from http://www.ncrel.org/engauge/

Factor Analysis Using SAS PROC FACTOR (1995). Retrieved January 7, 2005 from http://www.utexas.edu/cc/docs/stat53.html

Fink, A., ed. (1995). *The survey handbook*. Thousand Oaks, CA: Sage.

Goldstein, G. & Hersen, M. (eds.). (1984). *Handbook of Psychological Assessment.* New York, NY: Pergamon Press.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Guttman, L. (1950). The basis for scalogram analysis. In Stouffer et al. *Measurement & Prediction, The American Soldier*, Vol. 4. New York, NY: Wiley.

Linacre, J. M. (2000). Guttman Coefficients and Rasch data. *Rasch Measurement Transactions*, 14(2), 746-7.

Jaeger, R. M. (1990). *Statistics: A Spectator Sport.* Newbury Park, CA.: Sage Publications.

Johnson, D. (1997). The CODE 77 Rubrics. Retrieved February 19, 2004 from http://www.doug-johnson.com/dougwri/Rubbeg.HTM.

Johnson, D. (1999). *The indispensable teacher's guide to computer skills.* Worthington, Ohio: Linworth Publishing Inc.

Kerlinger, F. N. (1979) *Behavioral Research: A conceptual approach*. New York, NY: Holt, Rinehart & Winston.

Kinzer, C.K., & McKenna, M. C. (1999). *Using technology in your classroom literacy program: Current and future possibilities.* Paper presented at the 44th Annual Convention of the International Reading Association, San Diego, CA.

Knezek, G. & Christensen, R. (1998). *Internal Consistency Reliability for the Teachers' Attitudes Toward Information Technology Questionnaire*. Presented at the Society of Information Technology & Teacher Education (SITE)'s Ninth International Conference, Washington, DC, March 13, 1998.

McKenzie, J. (1994). From technology refusal to technology acceptance: a reprise. From *Now On: The Educational Technology Journal*, 4(9), 1-12. Retrieved October 28, 2004 from http://optin.iserver.net/fromnow/ may 94fno.html

McKenzie, J. (1998). Assessing professional growth in skill and use. *Staffdevelop.org*. Retrieved February 17, 2004, from http://staffdevelop.org/gauge.html

Moersch, C. (2004). Validity and Reliability. Retrieved January 5, 2005 from http://www.loticonnection.com/validandreliable.html

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw-Hill.

NCREL. (1997). Learning with Technology Profile Tool. Retrieved January 5, 2005 from http://www.ncrtec.org/capacity/profile/profwww.htm

Palmquist, M (2005). Internal Consistency. Retrieved January 5, 2005 from http://writing.colostate.edu/references/research/relval/com2a4.cfm.

Pawloski, R. W. (2003). Teacher candidates' perceptions of their capacity to meet national technology standards and the relationship to institutional preparation to teach with technology. *Dissertation Abstracts International*, 64(7), 2334. (UMI No. 3049366)

Russ-Eft, D. F. (1980). Validity and reliability in survey research. *American Institutes for Research in the Behavioral Sciences August*, 227, 151.

SEIR*TEC. (2000) Technology Integration Progress Gauge Profiler Version. Retrieved January 5, 2004 from www.seirtec.org/publications/ProgressGauge2000.pdf

TAGLIT. Retrieved January 5, 2004 from http://www.taglit.org/taglit/login.aspx

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

Trochim, W. M. (2002). Reliability and validity. Retrieved January 7, 2005 from http://www.socialresearchmethods.net/kb/rel&val.htm

Wepner, S. B., & Tao, L. (2002). From master teacher to master novice: Shifting responsibilities in technology-infused classrooms. *The Reading Teacher*, 55(7), 2-12.

Wilson, M. (Ed.) (1992-1994) *Objective Measurement: Theory into Practice* (Vols. 1-2). Norwood, NJ: Ablex.