DOCUMENT RESUME

ED 357 053 TM 019 771

AUTHOR	Taylor, Dianne L.; Campbell, Kathleen T.
TITLE	An Application-Based Discussion of Construct Validity and Internal Consistency Reliability.
PUB DATE	Jan 92
NOTE	26p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January 31-February 2, 1992).
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Construct Validity; Elementary School Teachers; Elementary Secondary Education; Heuristics; *Participative Decision Making; Research Methodology; *Research Utilization; Secondary School Teachers; *Teacher Participation; Test Reliability
IDENTIFIERS	*Internal Consistency

ABSTRACT

Several techniques for conducting studies of measurement integrity are explained and illustrated using a heuristic data set from a study of teachers' participation in decision making (D. L. Taylor, 1991). The sample consisted of 637 teachers. It is emphasized that validity and reliability are characteristics of data, and do not inure to tests as such (B. Thompson, 1991, 1992). An instrument found to produce valid, reliable data in one study may not do so in another because of phenomena external to the instrument. Because of this dilemma, an investigation of measurement integrity should be included as an important part of every study. Failure to do so may result in inaccurate findings that wrongly influence a field. Two tables present the illustrative data. (SLD)

An Application-based Discussion of Construct Validity and Internal Consistency Reliability

Dianne L. Taylor
Tulane University

Kathleen T. Campbell
University of New Orleans

Running Head: Validity and Reliability

Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, January 30, 1992.

Abstract

Several techniques for conducting studies of measurement integrity are explained and illustrated using a heuristic data set. It is emphasized that validity and reliability are characteristics of data, and do not inure to tests as such (Thompson, 1991, 1992). An instrument found to produce valid, reliable data in one study may not do so in another because of phenomena external to the instrument. Because of this, an investigation of measurement integrity should be included as an important part of every study. Failure to do so may result in reporting inaccurate findings which wrongly influence a field.

Measurement integrity is usually thought of in terms of validity and reliability, both of which are important concepts in the social sciences. Validity and reliability are attributes of data, and not of measures, notwithstanding popular misconceptions to the contrary (Thompson, 1991, 1992). These characteristics have important implications for the confidence with which inferences can be made in research.

Social science research often seeks to explore phenomena which cannot be directly observed or measured. Therefore, researchers assume that observable behaviors which can be measured are manifestations of the unobservable phenomena of actual interest in most research. This assumption must be tested. According to Carmines and Zeller (1979), the extent to which useful inferences can be made from data is dependent upon the relationship between an observable behavior which is measured and an unobservable phenomenon which cannot be measured. If the relationship is strong, one has greater confidence in interpreting the data; if it is weak, inaccurate and misleading conclusions are likely to result.

Instruments such as questionnaires and tests are, themselves, neither valid nor reliable, though these terms are often applied to measuring instruments (Thompson, 1991) rather than to data. Understanding the distinction is important. Misuse of an instrument may produce invalid data. In such a case, the problem would not be with the instrument, but with the application. Similarly, an instrument may measure a concept differently with different populations and in different settings. If an instrument is used under circumstances for which it was not designed, data are likely to be unreliable (Nunnally, 1967). As a consequence, it behooves the researcher to include tests of validity and reliability in every study. The current paper discusses measurement integrity in general, with emphasis on construct validity and internal consistency reliability. Applications of both are provided.

Overview of Validity and Reliability

Validity and reliability are not all-or-none propositions; for both these are matters of degree (Nunnally, 1967). In general, validity is the extent to which an instrument measures what it purports to measure (Nunnally, 1967; Popham, 1990). Reliability, on the other hand, is concerned with the consistency of the measurement in repeated applications (Kubiszyn & Borich, 1984; Nunnally, 1967). The two concepts are closely related. Reliability is a necessary, but insufficient, condition for validity (Lyman, 1971; Popham, 1990). Hence, evidence of reliability may be found when there is little evidence of validity; the reverse, however, cannot occur.

Because there is always error in measurement, no instrument produces data that are perfectly valid or reliable. Carmines and Zeller (1979) note that two types of error affect empirical data, random error and nonrandom error. Nonrandom error is systematic and affects the validity of a measurement (Carmines & Zeller, 1979;

Lyman, 1971). The greater the nonrandom error, the lower the validity.

Conversely, random error is unsystematic, varying somewhat with each measurement. Random error affects the reliability of a measurement; the greater the random error, the less reliable the measurement (Carmines & Zeller, 1979; Lyman, 1971). Random error is largely attributable to differences in subjects. However, sample size affects the degree of error present. As sample size increases, error is minimized (Nunnally, 1967).

Although validity and reliability are related concepts, they concern different aspects of measurement integrity. As might be expected, the procedures for assessing them are different. Therefore, each will be treated separately, with examples of techniques for investigating construct validity and internal consistency reliability provided.

Validity

Validity has been singled out by Lyman (1971) and Popham (1990) as the most important attribute in measurement. An instrument that does not measure what it is supposed to measure is of limited use. Because many concepts of interest in education are abstract and difficult to define, determining how to measure them is problematic. As a result, an instrument may "look valid, but measure something entirely different than what is intended" (Kubiszyn & Borich, 1984, p. 236). For example, a test that is supposed to measure intelligence, may measure guessing or achievement. Such an instrument would not be useful for making inferences about a person's intellectual ability.

Nunnally (1967) maintains that "the degree to which it is necessary and difficult to validate a measure of psychological variables is proportional to the degree to which the variable is concrete or abstract" (p. 84). Because many variables in the social sciences are abstract, Nunnally (1967) cautions that most instruments "should be kept under constant surveillance to see if they are behaving as they should" (p. 75).

There are several forms of validity. Among these are intrinsic, extrinsic, internal, external, divergent, convergent, face, ecological, content, criterion-related (both predictive and concurrent), and construct validity. The present paper will focus on content validity, criterion-related validity, and construct validity. The latter will also be demonstrated with examples.

Content validity concerns the extent to which the items on an instrument "are representative of some defined universe or domain of content" (American Psychological Association [APA], 1985, p. 10). Although establishing content validity involves the judgment of experts, rigorous steps are involved. First, literature is reviewed so that the universe of content applicable to the domain of interest can be specified (Carmines & Zeller, 1979). Second, items are produced which ensure both that the domain is adequately sampled and that, collectively, the items are representative of the important aspects of the domain (Popham, 1990). Third, the best items are selected. Review by experts can occur with each of these steps or at the conclusion when the final product has been produced (Popham, 1990).

Despite this demanding process, there is an "absence of... generally accepted procedures for assembling content-related evidence of validity" (Popham, 1990, p. 99), which has led some psychometricians to reject the concept altogether (Carmines & Zeller, 1979). Nonetheless, content validity is important, particularly in testing. Popham (1990) suggests that "when documenting the role of...experts in identifying the content for [an instrument], ...developers should indicate <u>how many</u> experts were selected from what sort of original list of invitees, <u>how many</u> opportunities those experts had to review the proposed content..., <u>what proportion</u> of the experts supported each content category...and so on" (p. 101, his emphasis).

A second commonly used validation method is criterion-related validity. Criterion-related validity is important when an instrument is being used to estimate a target behavior (the criterion) that is external to the instrument, such as when SAT scores are used to estimate success in college. There are two types of criterion-related validity, predictive and concurrent. The distinction between the two has to do with when the collection of data for the criterion variable occurs. For example, if scores on a test that is supposed to estimate success in school were correlated with examinees' grade point average (the criterion) at the time the test was administered, then evidence of concurrent validity would be obtained. If, however, the test were to estimate future performance and scores were correlated with data obtained at the future point, then predictive validity would be at issue. The SAT example above is a case in point. Carmines and Zeller (1979), Guilford (1954), and Popham (1990) all caution that when trying to establish criterion-related validity, the selection of an appropriate criterion is important

Evidence of criterion-related validity rests not with the judgment of experts, as was the case with content validity, but is obtained statistically through correlational procedures. The resulting validity coefficient ranges from .00 to 1.0 and represents the degree of relationship between the instrument and the criterion variable. The stronger the correlation, the stronger the evidence of criterion-related validity. Because complexity is inherent in both the instrument and the criterion variable, a high degree of correspondence between the two is precluded. Nunnally (1967) notes that only a modest correlation (r=.40 or r=.50) should be expected in many situations.

A third important form of validity is construct validity. A construct is a "theoretical construction about the nature of human behavior" (APA, 1985). Kubiszyn and Borich (1984) note that construct validity has been substantiated if there is correspondence with theory. Carmines and Zeller (1978) elaborate, describing construct validity as "concerned with the extent to which a particular measure relates to other measures consistent with theoretically derived hypotheses concerning the concepts (or constructs) that are being measured" (p. 23).

As Nunnally (1967) describes it, "a variable is literally a construct... [if] it is something that the scientist puts together from his own imagination, something that does not exist as an isolated, observable dimension of behavior" (p. 85). The speed of a runner, for example, is directly observable and measurable; speed is not theoretical and is not a construct. The runner's 'will to win,' however, is not directly observable or measurable; it is theoretical and is a construct. Other examples of constructs are intelligence, reading readiness, stress, love, creativity, and so on.

Several techniques can be used for measuring construct validity. Among these are Pearson correlation (Carmines & Zeller, 1979; Popham, 1990); canonical correlation (c.f., Halpin, Ralph, & Halpin, 1987; Sapp, Buckhalt, & Mitchell, 1985), and factor analytic procedures (Nunnally, 1967; see also, Thompson & Borrello, 1986; Tucker, 1990). Nunnally (1967) emphasizes that factor analysis is particularly important to establishing construct validity.

Because there are not criteria, nor a domain of content against which to test for construct validity, theoretical predictions must be developed that "lead directly to empirical tests involving measures of the [construct]" (Carmines & Zeller, 1979, p. 23). For example, one might theorize that individuals suffering from feelings of alienation would also be social isolates. To test for construct validity, a measure of alienation might be correlated with a measure of social isolation.

Carmines and Zeller (1979) describe three steps in construct validation.

> First, the theoretical relationship between the concepts themselves must be specified. Second, the empirical relationship between the measures of the concepts must be examined. Finally, the empirical evidence must be interpreted in terms of how it clarifies the construct validity of the particular measure. (p. 23)

If results of a validation study are negative, four possible causes should be explored (Carmines & Zeller, 1979, p. 25). First, it is possible that the instrument does not measure the phenomenon of interest and the researcher must return to the drawing board. Second, assumptions about the theoretically related variable may be incorrect. The 'related' variable may, in fact, not be related as predicted. Third, if the assumptions about the related variable are correct, the measurement of that variable may be problematic. And, fourth, the statistical procedures used to test the hypothesis of construct validity may have been improperly used.

Construct validation is complicated and may not yield satisfactory answers. According to Nunnally (1967), even after a meticulous investigation, "it is not possible to prove that any collection of observable [variables] measures a construct" (p. 97). Carmines and Zeller (1979) support this contention, warning that a single supporting study does not assure construct validity; rather, such evidence must be consistently found by numerous researchers across numerous studies.

An example of construct validation is provided using data from a study of teachers' participation in decision making (Taylor, 1991). The sample consisted of 637 regular education teachers who returned a questionnaire on decisional participation developed by Bacharach, Bauer, and Shedd (1986). The data were analyzed using principal components analysis in which the extracted structure was rotated to the varimax criterion. Four factors with associated eigenvalues greater than one prior to rotation (Thompson, 1989) were extracted. These factors included (a) an associated technology dimension, which involved decisions indirectly related to students and teacher, such as testing policies and teacher evaluation; (b) a managerial domain, tapping decisions that were rather remote from the classroom, such as facilities design; (c) a core technology II area that included matters related to textbooks; and (d) a core technology I dimension involving issues of what and how to teach. All questionnaire items had structure coefficients greater than .35 on at least one factor. Table 1 presents the items and their factor structure coefficients.

INSERT TABLE 1 ABOUT HERE.

The construct validity of the data was explored using a technique demonstrated by Thompson and Pitts (1981/82) and Thompson and Borrello (1986). Based on a program developed by Veldman (1967), the procedure determines "how closely obtained factors correspond with theoretically expected factors" (Thompson &

Borrello, 1986, p. 748). Thompson and Pitts (1981/82) explain the procedure.

> The calculated factors can be rotated to a position of "best fit" with a theoretically derived target matrix. The target matrix delineates how many factors are expected and the expected correlation between each item and each factor. The cosines of the angles between the hypothetical and the actual measures can be interpreted as validity (or correlation) coefficients. (p. 101)

The theoretically derived matrix for the present paper was established using related previous factor analytic work by Bacharach, Bamberger, Conley and Bauer (1990) in a study involving 1,531 teachers. Bacharach et al. also extracted four factors from the decisional participation questionnaire using a principal components analysis with a varimax rotation. The theoretically derived matrix for the current study consisted of 1s and 0s to reflect the structure matrix obtained in the Bacharach et al. (1990) study. Where an item in the Bacharach et al. (1990) study had a structure coefficient of at least .30 on a factor, it was assigned a theoretical structure coefficient of 1 in the target matrix, suggesting a perfect relationship with the factor, and a structure coefficient of 0 for the other three factors, suggesting no relationship with those factors. The second matrix was comprised of the rotated structure coefficients from the Taylor (1991) study.

Veldman's (1967) program takes two sets of orthogonal (or uncorrelated) factors, rotates them to "best fit" with each other, and computes cosines of the angles between the actual and theoretical matrices (Tucker, 1990). Because the factors are standardized to unit length, these cosines are also correlation coefficients (Gorsuch, 1983, pp. 62-66), and in this application can be thought of as validity coefficients.

Before the cosines among the factor axes are examined, however, relationships between the variables in the two sets must be considered (Gorsuch, 1983). According to Gorsuch (1983, p. 284) "if the mean cosine is low, it is not possible to relate factors because the relationships among the variables are different." Thompson and Pitts (1981/82) advise that "an item should have a cosine of roughly .80 or higher to be considered acceptable" (p. 103). Cosines among the variables in the current study fall between .52 and .95. Seven of the variables fail to meet the standard. The mean cosine of .81 is also somewhat low, but is strong enough to warrant further examination of results. The actual and theoretical matrices, and the cosines among the variables are presented in Table 2.

INSERT TABLE 2 ABOUT HERE.

Cosines between the actual and theoretical factors were .97, .89, .88, and .76, respectively, for factors one through four. The first three of these are in the acceptable range, thus providing evidence of construct validity; however, the cosine for factor four is below .80, suggesting that the factor is not an appropriate measure of teacher participation in decision making.

Establishing validity in a study indicates that evidence of reliability will also be found, since reliability is a necessary condition of validity. Three approaches to investigating reliability are presented below along with two examples of procedures for determining internal consistency reliability.

Reliability

Reliability was defined earlier as the consistency with which an instrument measures a variable in repeated applications. As noted, there are three general types of reliability. One type considers stability with repeated applications over time. A second approach compares different forms of an instrument. The third method "focuses on the consistency of [an instrument's] internal elements, namely, its...items" (Popham, 1990). All forms of reliability are based on correlational procedures and produce a reliability coefficient that ranges from .00 to 1.0.

It is generally held that reliability can be improved by increasing the number of items on an instrument. For example, Nunnally (1967) provides a discussion and formula for determining the number of items estimated to be needed to obtain a desired reliability. In the case of internal consistency reliability, however, Carmines and Zeller (1979) caution that because the reliability estimate is based on interitem correlations, increasing instrument length will improve the estimate only if the additional items do not diminish the average interitem correlation.

The first approach to reliability, the test-retest method, is used to determine if an instrument produces stable results in repeated administrations. The test-retest approach is conceptually simple, though it can be expensive and difficult to carry out in a practical sense. To assess test-retest reliability, an instrument is administered to a sample of subjects; a period of time is allowed to elapse, roughly two to four weeks; and the instrument is readministered to the same sample of subjects or to a subsample of the subjects. Scores from the two administrations are correlated. The average correlation among the items is the estimate of reliability.

Aside from the practical constraints to this approach, a measurement problem exists also. The prior experience of subjects with the test will influence the results of the second administration (Carmines & Zeller, 1979). If a greater amount of time is allowed to transpire to minimize this concern, problems of maturation and history will affect the results. Other forms of reliability are more useful.

A second type of reliability is called equivalent, or parallel, forms. Often in testing situations, it is necessary to have several forms of an instrument. For example, admission tests such as the GRE or the ACT may be taken more than once by prospective students trying to improve their scores. Equivalent forms of the test are needed so that memory of items from previous testing does not bias the results. Similar to the procedure described above, gathering evidence in a study of equivalent forms reliability involves administering a test to a group of subjects, allowing two to three weeks to elapse before the equivalent form of the test is administered, and correlating the scores from the two forms. Again, the correlation between the two is the reliability coefficient.

Another approach to reliability to examine the internal consistency of an instrument. Internal consistency is based on the "average correlation among items within a test" (Nunnally, 1967). Nunnally suggests that if the internal consistency of a new instrument is low, the instrument should be reconsidered.

Internal consistency can be determined several ways. If an instrument is to be dichotomously scored, for example, responses are scored as right or wrong, either of two Kuder-Richardson formulas are appropriate, *KR*20 and *KR*21. The *KR*20 formula, though computationally much simpler, is somewhat less accurate. Popham (1990) maintains that any loss in accuracy is more than offset by ease gained in computation. With computer access, however, any problem with difficult computations disappears. Popham also notes that the Kuder-Richardson formulas "can be thought of as representing the average correlation obtained from all possible split-half reliability estimates" (p. 133).

Split-half reliability estimates also fall within the internal consistency family. Split-half reliability estimates are derived by splitting the items on an instrument into two parts and correlating scores on both the parts. Because the resulting correlation is actually the reliability for each half of the instrument, a correction is needed which can be made using the Spearman-Brown prophecy formula (Carmines & Zeller, 1979).

A liability of this method is that the correlation between the two halves will differ somewhat depending on which items compose each half. Obtaining different correlations in a test of reliability means, by definition, that different reliability coefficients will be obtained. Differences in resulting coefficients were explored using the decisional participation questionnaire.

Items were first split odd-even. The split-half reliability estimate after the Spearman-Brown correction was a high .94. When the items were split 50-50, so that items 1 - 10 were in one half and items 11 - 19 in the other half, the reliability coefficient dropped to .81, still respectable, but much lower than previously obtained.

There are numerous ways to apportion items to compute a split-halves reliability estimate. Carmines and Zeller (1979) note that for "a 10-item scale, there are 125 different possible splits" (p. 43), many of which will yield slightly to substantially different reliability coefficients.

Cronbach's alpha, also referred to as coefficient alpha, is one of the most widely used measures of internal consistency. When items are scored dichotomously, alpha is equivalent to the KR21 result, but alpha can also be computed for nondichotomously-scored items. The computations are based on the number of items and the mean interitem correlation. Therefore, the value of alpha can be increased by increasing the number of items (although there is a point of diminishing returns with this approach) and/or by increasing the intercorrelations among the items. Increasing items will not improve alpha, however, if the additional items decrease the interitem correlation (Carmines & Zeller, 1979).

Coefficient alpha was computed for the decisional participation instrument and for each of the four factors. For the total scale, alpha was .90; for each factor alpha was .84, .78, .89 and .66, respectively. Although these coefficients are sufficiently strong to warrant some confidence in the reliability of the data, the .66 estimate for factor four is borderline. It will be remembered this same factor was not found to be a valid measure of participation in the validation study. Clearly, there is a measurement integrity problem with the fourth factor.

Summary

Several techniques for conducting studies of measurement integrity have been explained. Examples and interpretations of measures of construct validity and internal consistency reliability have been provided. From these examples, it is obvious that establishing measurement integrity is wrought with the same difficulties as other aspects of statistical analysis and interpretation.

It should be reiterated in closing that validity and reliability refer to characteristics of data, and do not inure to tests as such (Thompson, 1991, 1992). An instrument found to produce valid, reliable data in one study may not do so in another because of phenomena external to the instrument. Therefore, an investigation of measurement integrity should be included as an important part of every study. Failure to do so may result in reporting inaccurate findings which wrongly influence a field.

References

American Psychological Association (1985). Standards for educational and psychological testing. Washington, DC: Author.

Bacharach, S. B., Bamberger, P., Conley, S. C., & Bauer, S. (1990). The dimensionality of decision participation in educational organizations: The value of a multi-domain evaluative approach. Educational Administration Quarterly, 26, 126-167.

Bacharach, S. B., Bauer, S. C., & Shedd, J. B. (1986). The learning workplace: The conditions and resources of teaching. (ERIC Document Reproduction Service No. ED 279 614)

Carmines, E. G. & Zeller, R. A. (1979). Reliability and validity assessment. Beverly Hills: Sage.

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

Guilford, J. P. (1954). Psychometric methods (2nd ed.). New York: McGraw-Hill.

Halpin, G., Ralph, J., & Halpin, S. (1987, November). a study of the validity and reliability of the Adult Career Concerns Inventory. Paper presented a the annual meeting of the Mid-South Educational Research Association, Mobile, AL. (ERIC Document Reproduction Service No. ED 291 767)

Kubiszyn, T. & Borich, G. (1984). Educational testing and measurement. Glenview, IL: Scott, Foresman.

Lyman, H. B. (1971) Test scores and what they mean (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Nunnally, J. C. (1967). Psychometric theory. New York: McGraw-Hill.

Popham, W. J. (1990). Modern educational measurement: A practitioner's perspective (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Sapp, G. L., Buckhalt, J. A., & Mitchell, M. (1985). A comparison of the K-ABC and WISC-R: A validation study. Paper presented at the annual meeting of the National Association of School Psychologists, Las Vegas, NV. (ERIC Document Reproduction No. ED 285 892)

Taylor, D. L. (1991). The effect of teacher participation in decision making on selected teacher and student outcomes. Unpublished doctoral dissertation, University of New Orleans, New Orleans.

Thompson, B. (1989). Prerotation and postrotation eigenvalues shouldn't be confused: A reminder. Measurement and Evaluation in Counseling and Development, 22(3), 114-116.

Thompson, B. (1991). Review of Generalizability theory: A primer by R.J. Shavelson & N.W. Webb. Educational and Psychological Measurement, 51, 1069-1075.

Thompson, B. (1992). Two and one-half decades of leadership in measurement and evaluation. Journal of Counseling and Development, 70, 434-438.

Thompson, B., & Borrello, G. (1986). Construct validity of the Myers-Briggs Type Indicator. Educational and Psychological Measurement, 46, 745-752.

Thompson, B. & Pitts. M. C. (1981/82). The use of factor adequacy coefficients. Journal of Experimental Education, 50, 101-104.

Tucker, M. L. (1990). Higher education leadership: Transformational leadership as a predictor of satisfaction, effectiveness, and extra effort. Unpublished doctoral dissertation, University of New Orleans, New Orleans.

Veldman, D. J. (1967). Fortran programming for the behavioral sciences. New York: Holt, Rinehart, and Winston.

Table 1
Rotated Factor Matrix for Actual Decisional Participation

Item	Factors I	II	III	IV	h^2
Associated technology					
Student rights	**.768**	.153	.167	.134	.659
Standardized testing policy	**.747**	.191	.145	-.027	.612
Student discipline codes	**.701**	.281	.147	-.045	.594
Reporting student achievement	**.658**	.092	.059	.328	.553
Grading policies	**.647**	.107	.134	.294	.534
Teacher's performance evaluation	**.527**	.302	.102	.139	.399
Staff development	**.444**	.326	**.427**	.006	.497
Students' assignment to class	**.423**	.221	.019	.233	.282
Managerial					
Budget development	.289	**.764**	.168	-.043	.697
Spending priorities	.260	**.755**	.199	-.068	.681
Staff hiring	.236	**.661**	-.001	.060	.497
Teacher's assignment to school	.025	**.626**	.030	.274	.468
Designing facilities	**.457**	**.481**	.173	.131	.487
Removal for special instruction	.337	**.377**	.066	.204	.302
Core technology II					
Texts/workbooks available	.149	.123	**.874**	.272	.875
Texts/workbooks used	.220	.136	**.870**	.214	.870
Core technology I					
How to teach	.110	.072	.189	**.794**	.683
What to teach	.283	.016	.164	**.767**	.693
Teacher's subject/grade assignment	.105	.422	.165	**.475**	.442
Eigenvalue	3.940	2.975	1.997	1.995	

Table 2

Matrices for Testing Construct Validity

Item	Matrix A Theoretically derived target				Matrix B Actual structure coefficients				Cosines
1	0	0	0	1	.03	.63	.03	.27	.69
2	0	0	0	1	.11	.42	.17	.48	.70
3	1	0	0	0	.42	.22	.02	.23	.90
4	1	0	0	0	.34	.38	.07	.20	.73
5	0	0	1	0	.46	.48	.17	.13	.60
6	0	0	1	0	.29	.76	.17	-.04	.88
7	0	0	1	0	.26	.76	.20	-.07	.90
8	0	0	1	0	.24	.66	-.00	.06	.80
9	1	0	0	0	.53	.30	.10	.14	.89
10	1	0	0	0	.70	.28	.15	-.05	.89
11	1	0	0	0	.75	.18	.15	-.03	.92
12	1	0	0	0	.65	.12	.13	.29	.94
13	1	0	0	0	.66	.09	.06	.33	.95
14	1	0	0	0	.77	.15	.17	.13	.95
15	0	1	0	0	.28	.02	.16	.77	.60
16	0	1	0	0	.11	.07	.19	.79	.65
17	0	1	0	0	.22	.14	.87	.21	.92
18	0	1	0	0	.15	.12	.87	.27	.96
19	0	1	0	0	.44	.33	.43	.01	.52