Research in Nursing & Health, 1991, 14, 315–320

Focus on Psychometrics The Multitrait–Multimethod Approach to Construct Validity

Sandra L. Ferketich, Aurelio J. Figueredo, and Thomas R. Knapp

The multitrait–multimethod matrix approach as proposed by Campbell and Fiske (1959) was an important contribution to our understanding of the nature of validation procedures. There are, however, problems encountered when using the Campbell and Fiske (1959) approach. The purpose of this article is to discuss the method and selected problems, and to propose an alternate approach to address those problems.

Campbell and Fiske (1959) provided important insights into the nature of validation procedures. In particular, they proposed the multitrait–multimethod matrix (MTMM matrix) as a way to assess convergent and discriminant validity as well as to estimate the effect of method variance on validity assessments. However, there are problems with the traditional bivariate analysis of the MTMM matrix that may severely limit the utility of the method. The purpose of this article is to discuss the method and some of the associated problems and to suggest ways in which these problems might be addressed. In a second article, to be published in the next issue, we will provide an example of such an approach illustrating an extension of the method. Specifically, we will explore the use of confirmatory factor analysis (CFA) in order to assess convergent and discriminant validity.

Campbell and Fiske Approach

Independent measures of traits are essential information for assessing construct validity, whether the measures are taken at the same or different times. The basic underlying tenets of the Campbell and Fiske MTMM matrix approach (1959) are that (a) tests designed to measure the same construct should correlate highly among themselves and (b) tests measuring one construct should not correlate with tests measuring other constructs. Thus, based on the first tenet, convergent validity is supported by the presence of relatively strong correlations among measures of the same construct; and based on the second tenet, discriminant validity is supported by the presence of relatively small correlations among tests measuring other constructs regardless of the method used.

The MTMM matrix consists of sets of intercorrelations that are evaluated according to four criteria as described by Campbell and Fiske (1959). Table 1 illustrates the general form of the matrix. Using general conventions, we have designated three methods as 1, 2, and 3 and three traits as A, B, and C for illustrative purposes. The blocks along the diagonal are termed *monomethod blocks* and are outlined with a solid line. Each value in each monomethod block represents a correlation between traits measured by the same method. Typically, the main diagonal within each monomethed block contains reliability estimates. Since the upper triangle of these blocks is composed of redundant correlations among traits, the upper triangle is usually omitted from the matrix. The blocks below the main diagonal of monomethod blocks are termed heteromethod blocks and are outlined by a solid and a broken line. The blocks contain correlations among each of the traits mea-

This is the fifth article in the series, "Focus on Psychometrics," edited or contributed by Sandra Ferketich, PhD, RN. Dr. Ferketich is an associate professor in the College of Nursing and the Director of the Predoctoral and Postdoctoral Clinical Nursing Research Instrumentation Fellowship Program (NRSA Institutional Grant No. 5T32 NR07029) at the University of Arizona. Aurelio J. Figueredo, PhD, is an assistant professor in the Department of Psychology, University of Arizona. Thomas R. Knapp, EdD, is a professor in the College of Nursing and the College of Education, Ohio State University.

Requests for reprints can be addressed to Dr. Sandra Ferketich, College of Nursing, University of Arizona, Tuscon, AZ 85721.

Table 1. MTMM Matrix for 3 Traits (A, B, C) and 3 Methods (1, 2, 3)

	Method 1			Method 2			Method 3		
	Trait A	B	C	Trait A	B	C	Trait A	B	C
Method 1									
Trait A	$r_{A_1 A_1}$								
Trait B	$r_{A_1 B_1}$	$r_{B_1 B_1}$							
Trait C	$r_{A_1 C_1}$	$r_{B_1 C_1}$	$r_{B_1 C_1}$						
Method 2				MONOMETHOD BLOCKS					
Trait A	$\underline{r_{A_1 A_2}}$ · · · ·	$r_{A_2 B_1}$ · · · ·	$r_{A_2 C_1}$						
Trait B	$r_{A_1 B_2}$	$\underline{r_{B_1 B_2}}$	$r_{B_2 C_1}$						
Trait C	$r_{A_1 C_2}$	$r_{B_1 C_2}$	$\underline{r_{C_1 C_2}}$						
Method 3	HETEROMETHOD BLOCKS								
Trait A				$\underline{r_{A_2 A_3}}$	$r_{A_3 B_2}$	$r_{A_3 C_2}$	$r_{A_3 A_3}$		
Trait B				$r_{A_2 B_3}$	$\underline{r_{B_2 B_3}}$	$r_{B_3 C_2}$	$r_{A_3 B_3}$	$r_{B_3 B_3}$	
Trait C				$r_{A_2 C_3}$	$r_{B_2 C_3}$	$\underline{r_{C_2 C_3}}$	$r_{A_3 C_3}$	$r_{B_3 C_3}$	$r_{C_3 C_3}$

sured by pairs of the methods. Since there are three methods in our example, there are three possible combinations of methods, Method 1 with 2, Method 1 with 3, and Method 2 with 3. The main diagonals in the heteromethod blocks contain correlations for the same trait measured by different methods and are underlined. These values are part of the validity diagonals and each element represents a monotrait–heteromethod estimate. Thus, each value assesses convergent validity, and is a direct assessment of the same trait measured by independent methods.

Campbell and Fiske (1959) describe four conditions that should be taken into account when examining a MTMM matrix. These conditions are listed below.

1. The correlations in the validity diagonal(s) should be of a sufficient magnitude to encourage further exploration of validity and, of course, should be significantly different from zero.
2. Each of the correlations in the validity diagonal(s) should be higher than the values lying in the specific correlation's column and row. That is, the specific correlation should be higher when the same trait, different methods are correlated than when that trait is correlated with a different trait, different methods. For illustration, in Table 1 the uppermost heteromethod block has a dotted line to connect an individual correlation with those in its column and row.
3. The correlation between two different methods of measurement of the same trait should be higher than correlations between that trait and another trait even though the methods of measurement are the same.
4. The pattern of interrelationships between traits should be similar in the heterotrait triangles, correlations between traits, whether in the monomethod or heteromethod blocks.

As can be seen, the researcher is required to make a number of judgment calls regarding the interpretation of values within the MTMM matrix. A thoughtful approach to interpretation of the values within the matrix is an asset. However, problems may still arise in interpretation due to inherent problems with the MTMM approach.

Selected Problems with Interpreting the MTMM Matrix

There are three basic issues that we want to bring to the reader's attention that have created problems in the past in interpreting the results of a convergent and discriminant validation study using the MTMM matrix approach. The first issue concerns the criteria for evaluating the magnitude of similarities and differences among the interrelationships of the various components of the MTMM matrix. The second is a substantive issue when selecting specific traits and methods to be included in the study. Third, each "trait–method unit" (Campbell & Fiske, 1959) is composed of trait, method, and random variance. Using the methods proposed by Campbell and Fiske (1959), it is not possible to analyze and separately estimate these components. Thus, if the assumptions of the method in regard to the nature of the error components as proposed by Campbell and Fiske (1959) are not met, it is difficult to interpret the results of the validation study. See Schmitt and Stults (1986) for a more extensive discussion.

Criteria for Evaluating the Magnitude of Similarities and Differences

Campbell and Fiske (1959) did not determine specific criteria for assessing the magnitude of any similarities or differences. They challenge the researcher to assess individual correlations in the matrix as well as patterns in the matrix. The individual correlations must be statistically different from zero and substantively interesting. The latter is often in the eye of the beholder. In practice, consistent patterns frequently do not exist. The researcher is left trying to interpret what it means when some but not all of the criteria are met. What is probably equally troublesome is that reliability of the measures is an important issue that is not sufficiently addressed (Alwin, 1974). Correlations can be attenuated by poor instrument reliability. Thus, the magnitude of correlations and patterns of interrelationships may be unstable. It may be quite difficult to find two to three different reliable measures of two to three different traits for use in a MTMM matrix. Thus, the problems of lack of consistency in relationships and patterns, due to unreliability, can create difficulty in interpreting the results of a MTMM matrix approach.

Selection of Traits and Methods

The second issue of selection of traits and methods is more attributable to researcher choice. In applying MTMM to nursing research, investigators must carefully choose the discriminant trait(s) with which the trait of particular interest is to be contrasted. A discriminant trait must be a trait that is similar to, and easily confused with, the focal trait. For example, if one is using a MTMM approach to test the construct validity of an instrument that is alleged to be an operationalization of anxiety, a sensible choice for the discriminant trait would be fear, or stress; not height, mathematical aptitude, or comfort. Neither height nor mathematical aptitude can be substantively confused with anxiety, nor can comfort which may be defined as the opposite of anxiety. The nursing research literature contains an unfortunately large number of examples of supposed applications of MTMM where the discriminant trait is achievement, for intelligence; hopelessness, for hope; or the like.

The choice of method(s) is equally important. If one has two traits and two methods, the two methods must really *be different*. Methods such as self-report and observation are different; but long form and short form, multiple choice and true–false, and self-report at Time 1 and self-report at Time 2 are not different. Although it involves more than two traits measured by each of two methods, Polit and Hungler's (1991) example is a model of the appropriate choice of contrasting traits and methods.

A common problem with some applications of MTMM in the nursing research literature is that the investigators seem to want to "have it both ways." That is, they examine the relative magnitudes of the individual instrument reliabilities; the within-trait, between-method correlations; the between-trait, within-methods correlations; and the between-trait, between-method correlations. They may not obtain the results that they hoped they would get. For example, the between-trait, within-method correlations turn out to be higher than the within-trait, between-method correlations. The researchers, however, then go on to claim that the instrument of particular concern (a self-report measure of anxiety, for example) is valid since scores on that instrument correlate so "nicely" with the self-report measure of fear!

A caution regarding the MTMM: even if all of the reliabilities and intercorrelations come out "right," all you know is that your instrument and its alternative-method(s) counterpart(s) are converging on the same trait and can be discriminated from another trait. From the empirical evidence alone, you do not know whether they are converging on the trait that you allege to be measuring. That requires an additional act of faith, theoretical justification, or both.

Trait–Method Unit Issues

The following discussion is focused on an extension of the MTMM approach through the use of a CFA model. If you are unfamiliar with causal model and latent variable model language, see Ferketich and Verran (1990).

The central problem in the analysis of the MTMM matrix is an intentional consequence of its design. If every test is indeed a trait–method unit, then any bivariate correlation between such tests is, at least potentially, attributable to *common factor* variance derived from (a) shared *trait* variance, (b) shared *method* variance, or (c) some combination of both. In addition, such a bivariate correlation is, at least potentially, further limited by each test's *unique* variance, consisting of (d) residual test *specific* variance, not generally associated with the method, or (e) random *error* variance, i.e. test *unreliability*, or (f) some combination of both. Indeed, the entire MTMM approach is based on the overt realization of these limitations. Single bivariate correlations thus are not interpreted independently, but are systematically compared and contrasted to reveal the underlying patterns of association.

Although the MTMM approach made much conceptual use of hypothetical constructs such as traits and methods, which implied latent common factors, the basic unit of analysis remained at the level of bivariate correlations between tests representing theoretically specified interrelationships. The traditional methods of *exploratory* factor analysis (EFA) available at the time, however, were not adequate for MTMM analysis because the factor extraction procedures could not be constrained to distinguish between trait and method correlations in creating common factors. Uninterpretable combinations of trait and method variance could thus be produced opportunistically by the atheoretical or data-driven factor extraction procedures of traditional EFA.

A more sophisticated approach to this problem was made possible by Jöreskog's (1969) development of CFA. Widaman (1985) used CFA to construct common factors representing the latent traits and methods of the MTMM approach. By representing a convergent validity coefficient as a trait factor loading, trait variance was indicated by a correlation between the test and a trait construct. Convergent validity was thus not merely a correlation with some other equally fallible test that presumably measured the same trait, but directly with the underlying latent trait itself. Similarly, method variance was represented as a method factor loading indicated by a correlation between the test and a method construct, not merely a correlation with some other equally fallible test that shared the same general method, but with the reified latent method itself.

The CFA approach made for a cleaner operationalization of both trait and method common factor variance. For example, whereas previously there were different "convergent validity coefficients" derivable between any two tests purported to measure the same trait, the CFA approach provided a single factor loading coefficient indicating an average correlation with a sample of converging tests and, inferentially, with a hypothesized population of such tests. This CFA approach also greatly clarified the concept of *discriminant validity*. Previously, undesirable high correlations between tests that purported to measure different traits could be attributed to one of three possibilities: (a) there are high correlations between both tests and a common method factor, (b) there are high correlations between either (or both) of the tests and more than one trait factor, or (c) there are high correlations between the underlying trait factors themselves. There was no simple way to specify which single one or combination of these three possible causes was responsible for the lack of discriminant validity.

The *first possibility* implies a systematic bias in the common method. This limits full discrimination between traits by that particular method, but does not further specify any possible correlations between either (a) the latent traits and the two tests or (b) the underlying traits themselves. This first possibility is modeled in CFA as high and significant loadings from the same method factor on the two tests. The *second possibility* implies a condition of factorial complexity in either (or both) of the two tests, thus limiting discrimination between traits by the affected tests. It does not, however, further specify any possible correlations between the underlying traits themselves. This second possibility is modeled in CFA as multiple high and significant loadings from different trait factors on either (or both) of the tests. The *third possibility* implies a real statistical relationship between the underlying traits themselves, thus limiting full independence between latent trait constructs, regardless of the methods used to measure them. This third possibility is modeled in CFA as high and significant intercorrelations between the trait factors. These are three very distinct causal conditions that would all have been subsumed previously under the common rubric of *poor discriminant validity*, but tell completely different stories about why such poor discrimination is obtained.

In summary, the CFA approach makes fully explicit the conceptual relations that are only implicit in the traditional bivariate analysis of the MTMM matrix. The various traits and methods of the MTMM matrix are more tangible as common factors than as abstract organizing principles defining the critical regions of the bivariate correlation matrix and specifying the theoretical interrelationships between them. The direct contributions of these latent trait and method constructs to the test scores and intercorrelations are more clearly identified and quantified. For example, whereas the traditional MTMM analysis could do little more than help us detect problems in discriminant validity, the CFA approach permits us to identify their causes, given an adequate MTMM study, by providing more specific diagnoses of the nature of these problems.

On the other hand, we should not deceive ourselves into thinking that the CFA approach can somehow overcome the limiting assumptions of the basic MTMM design. CFA merely provides a more concise and explicit representation of the same system of relationships. As with any model, CFA is subject to the principle of *garbage in→garbage out* (GIGO). For example, if the methods are not truly heterogeneous or the traits are not meaningful constructs, the CFA approach will do little more than formalize our conceptual errors. CFA will therefore not magically correct any flaws in the design of the MTMM experiment. For those who are easily mesmerized or intimidated by the seeming authority of mathematical models, the elegance of the CFA representation can pose the hazard of suspending the exercise of critical judgment. However, since there is no statistical analysis that can be guaranteed foolproof without requiring some careful thought processing, we should not be deterred from judicious use of what is currently the most sophisticated analytical technique available for construct validation.

In the next article in this series, we will provide an example that is analyzed using the CFA approach. We will discuss how this solution differs from the traditional MTMM approach.

REFERENCES

Alwin, D.F. (1974). Approaches to the interpretation of relationships in the multitrait–multimethod matrix. In H. Costner (Ed.), *Sociological methodology: 1973–1974* (pp. 79–105). San Francisco: Jossey-Bass.

Campbell, D.T., & Fiske, D.W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Ferketich, S.L., & Verran, J.A. (1990). Causal modeling. In L. Moody (Ed.), *Advancing nursing* *science through research: Vol. II* (pp. 111–144). Newbury Park, CA: Sage.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183–202.

Polit, D.F., & Hungler, B.P. (1991). *Nursing research: Principles and methods* (4th ed.) (pp. 378–379). Philadelphia: J. B. Lippincott.

Schmitt, N., & Stults, D.M. (1986). Methodology review: Analysis of multitrait–multimethod matrices. *Applied Psychological Measurement, 10*, 1–22.

Widaman, K.F. (1985). Hierarchically nested covariance structure models for multitrait–multimethod data. *Applied Psychological Measurement, 9*, 1–26.