1982

**Multitrait-multimethod matrix assessment of selected neuropsychological instruments**

Valerie Kim Sweeney

Follow this and additional works at: [https://scholarworks.lib.csusb.edu/etd-project](https://scholarworks.lib.csusb.edu/etd-project)

Part of the [Experimental Analysis of Behavior Commons](https://scholarworks.lib.csusb.edu/etd-project)

**Recommended Citation**


This Thesis is brought to you for free and open access by the John M. Pfau Library at CSUSB ScholarWorks. It has been accepted for inclusion in Theses Digitization Project by an authorized administrator of CSUSB ScholarWorks. For more information, please contact scholarworks@csusb.edu.
MULTITRAIT-MULTIMETHOD MATRIX ASSESSMENT OF
SELECTED NEUROPSYCHOLOGICAL INSTRUMENTS

A Thesis
Presented to the
Faculty of
California State College
San Bernardino

In Partial Fulfillment of
the Requirements for the Degree
Master of Arts
in
Psychology

Valerie Kim Sweeney
June 1982
MULTITRAIT-MULTIMETHOD MATRIX ASSESSMENT OF
SELECTED NEUROPSYCHOLOGICAL INSTRUMENTS

A Thesis
Presented to the
Faculty of
California State College
San Bernardino

by
Valerie Kim Sweeney
June 1982

Approved by:
ABSTRACT

The multitrait-multimethod matrix method of assessing construct validity proposed by Campbell and Fiske (1959) was used to evaluate verbal and motor skills tests. Twenty-five children evidencing neuropsychological and/or psychological abnormalities were used as subjects. Each trait was assessed using three independent neuropsychological methods, a memory measure, a standardized performance measure, and a non-standardized performance measure. Verbal skills were evaluated using scores from the Rey Auditory-Verbal Learning Test (AVLT), the Verbal Wide Range Achievement Test (WRAT), and the Cleaves School Adjustment Scale, while scores from the Graham-Kendall Memory for Designs Test (MFD), the Developmental Test of Visual-Motor Integration (VMI), and The Cleaves Motor Index provided measures of motor skills. The results of this investigation provided evidence for both convergent and discriminate validity involving some, but not all, of the neuropsychological measures employed. To evaluate motor skills, the results encourage the use of the VMI and the MFD, while an assessment of verbal skills is encouraged using the Cleaves School Adjustment Scale and either the AVLT or the Verbal WRAT, but not for using both the AVLT and the Verbal WRAT. An extension of the study involving a larger number of subjects is recommended.
TABLE OF CONTENTS

Abstract ................................................................. iii
List of Tables ........................................................... v
Introduction ............................................................... 1
  Construct Validity ..................................................... 1
  Multitrait-Multimethod Analysis ................................. 2
  Research Problem .................................................... 5
Method ................................................................. 7
  Subjects ............................................................... 7
  Apparatus ............................................................ 7
  Procedure ........................................................... 11
Results and Discussion ................................................. 17
  Convergent Validity ................................................ 18
  Discriminate Validity .............................................. 18
General Discussion .................................................... 21
References .............................................................. 24
LIST OF TABLES

1. Hypothetical 3 X 3 Multitrait-Multimethod Matrix Showing Four Conceptual Units. 12

2. Social Intelligence and Mental Alertness Subtest Intercorrelations from Thorndike's Data. 16

3. Verbal and Motor Skills Intercorrelations 17
INTRODUCTION

Construct Validity

In all fields of psychology involving the measurement of individual differences, the validity and reliability of the measurement instruments used is of primary concern. Unless the theoretical construct of interest is accurately measured by the test given, construct validity is non-existent. Campbell and Fiske (1959) addressed the validation process and discussed four important relationships bearing on this process.

1. A trait is said to possess convergent validity to the extent that there is a high correlation between maximally independent methods designed to measure the same trait. Fiegl (1958) refers to this criterion as establishing construct validity by "triangulation in logical space" (p. 401).

The extent to which two or more methods are actually independent may be viewed as a matter of degree. Campbell and Fiske, for example, suggest that reliability, which is "the agreement between two efforts to measure the same trait through maximally similar methods," and validity, which is "represented in the agreement between two attempts to measure the same trait through maximally different methods" (p. 83) be seen as appearing on a continuum. The utilization of methods which are not entirely independent, however,
does not necessarily rule out a validity evaluation. It is advisable, however, to keep in mind that one is most likely to attain a measure of relative validity which includes some amount of shared method variance.

2. In addition to the more common confirmation of construct validity by convergent validation, another essential criterion is a measure of discriminate validity. Discriminate validation requires that there be a low correlation between tests designed to measure different traits.

3. Each measurement test or instrument must be regarded not merely as representative of a particular trait, but as a combination of the trait assessment and the method by which it is assessed. Campbell and Fiske term this a trait-method unit and indicate that a certain amount of systematic variance is inevitable. Hence, the scores obtained are invalidated to the extent that they are affected by this method variance. Such sources of invalidity have been variously termed "halo effects" (Thorndike, 1920), "test-form factors" (Vernon, 1958), and "response sets" (Cronbach, 1950).

4. Since construct validity is not associated with a particular correlation but with a pattern of correlations, individual validity measures do not indicate the extent to which a particular test actually measures what it was intended to measure.

Multitrait-Multimethod Analysis

In an effort to assess both convergent and discriminate
validity and to evaluate accurately the effects of trait and method variance, all of which are necessary to adequately establish construct validity, the use of a multitrait-multimethod matrix is advocated. This matrix utilizes data collected on test administrations for a number of subjects. In order to use the matrix properly, it must include all of the intercorrelations for at least two methods of assessment, for a minimum of two different traits. According to Campbell and Fiske, construct validity cannot be measured directly; construct validity is established by implication, reflecting an expected pattern in the results. The multitrait-multimethod matrix provides an opportunity to analyse this pattern.

Since the appearance of the Campbell and Fiske article, the number of researchers utilizing some form of the multitrait-multimethod matrix to evaluate test validity has become legion in many areas of psychology. Almost any combination of tests, inventories, and populations may be found. A survey of some of the recent literature amply illustrates the matrix's utility. Blaha, Fawaz, and Wallbrown (1979) examined the validity of the Bender Visual-Motor Gestalt Test (BVMG), Matching Familiar Figures Test, Draw-A-Person Test, and the Slosson Intelligence Test given to 74 Black, middle class first graders. The authors concluded from the matrix analysis that the clinical validity of the BVMG was not restricted to a single area such as general intelligence,
but was spread relatively evenly over all stages of information processing. Mellon and Crano (1977) used the matrix to examine three academic traits, assessed by standardized tests and teachers ratings, for 4,700 British schoolchildren. This study included a longitudinal data component which, in combination with the multitrait-multimethod approach, was said to provide a much more complete analysis of existing methods variance than other available techniques. Another example of the matrix's utility is an analysis of three self-concept inventories including the How I see Myself Scale, the Sears Self-Concept Inventory, and the Piers-Harris Children's Self-Concept Scale administered to 103 American third and sixth graders (Winne, Marx, & Taylor, 1977). The construct validity analysis provided by the multitrait-multimethod matrix enabled these researchers to isolate and compare individual facets of self-concept appearing under different labels on the separate inventories. Smith and Singer (1977) conducted a comparison of the Matching Familiar Figures Test, the Kansas Reflection-Impulsivity (R-I) Scale for Preschoolers and the Test of R-I in Social Content, given to 115 six and eleven year old Educable Mentally Retarded subjects. These researchers attempted to assess the reflection-impulsivity dimensions of time and error. The results indicated that the use of the matrix revealed generally weak construct validity, especially at the higher age level.
At the same time that the use of the multitrait-multimethod matrix has proliferated as a highly effective means of evaluating construct validity, the number of available measurement instruments has also increased dramatically. The field of neuropsychology, for example, which must depend almost exclusively upon various methods of diagnostic testing to evaluate and plan treatment for its clients, is no exception. In addition to using a large number of "tried-and-true" testing instruments—some without a detailed evaluation of their construct validity—many, if not most, private neuropsychologists are utilizing some type of "self-developed" subjective or objective scale to assist them in client diagnosis. Many of these self-developed instruments may possess construct validity which rivals, or even exceeds, that of current normative instruments now in widespread use.

In order for the practicing neuropsychologist to properly carry out his function, data regarding the construct validity of both available and newly developing measurement instruments must exist. Only with this kind of information, perhaps best provided by the multitrait-multimethod matrix, can an informed selection between existing testing instruments be made or can the construction of new measurement instruments proceed.

Research Problem

In an attempt to contribute to the literature with regard to the evaluation of the construct validity of neuropsycholog—
ical instruments, a 2 X 3 multitrait-multimethod matrix was constructed and analysed. This matrix involved two traits, verbal skills and motor skills, each assessed by three methods, a memory measure, a standardized performance measure, and a non-standardized performance measure. Verbal skills were measured by using scores from trials one through five of the Rey Auditory-Verbal Learning Test (AVLT) (memory measure), scores from the Reading and Spelling subtests of the Wide Range Achievement Test (Verbal WRAT) (standardized performance measure), and the Reading and Spelling components of the Cleaves School Adjustment Scale (non-standardized performance measure). The latter measure involves a subjective determination made by the psychologist and is based on the subject's school grades. Motor skills were measured by the score from the Graham-Kendall Memory for Designs Test (MPD) (memory measure), scores from the Developmental Test of Visual-Motor Integration (VMI) (standardized performance measure), and the Cleaves Motor Index (non-standardized performance measure). The latter is a factor analytically derived assessment of motor competence based on scores from the performance subtests of the Wechsler Intelligence Scale for Children - Revised (WISC-R). With the exception of the non-standardized performance measures which have been derived in response to local needs and which were considered as newly developed measurement instruments undergoing construct validity assessment, the testing instruments evaluated are generally accepted verbal and motor assessment measures.
METHOD

Subjects

The subjects were 25 children, 16 males and 9 females, ranging in chronological age from 7 to 16 years. All possess some type of neuropsychological and/or psychological abnormalities and were referred by state or local agencies to the Child Study Center, La Sierra, California, for testing and treatment. Some children also possess physical handicaps. All ethnic and socio-economic groups were represented, although the sample representation of minority and lower SES children was higher than in the general population. Most of the subjects attend public schools; all school performance levels were represented.

Apparatus

Testing instruments used in formulating the matrix were administered individually by a licensed psychologist at the Child Study Center.

The AVLT is a test of verbal learning and immediate memory span that detects cases of proactive and retroactive interference and confusion on memory tasks. It also measures retention following an interpolated task. The subject is presented with a list of 15 words read by the examiner. Following the list presentation the subject is to recall as
many words as possible. This constitutes one trial. The examiner then reads the list again and the subject again attempts total recall. Five verbal memory trials using the same words are given, followed by a list of 15 different words and a recall trial on this second list. A sixth trial on the original list and a recognition trial complete the test. The score for each trial is the number of words recalled correctly (Lezak, 1976, pp. 352-356). The sum of the scores on the first five trials was used as the verbal memory measure.

The WRAT is a performance test of academic competence and language processing. It is composed of three subtests, Reading, Spelling, and Arithmetic, with grade equivalents and percentages by age calculated from standard scores for each subtest. The test is available in two age ranges or "Levels": Level I is to be used for subjects 5 to 11 years of age, while Level II tests those aged 12 years to adult. The Spelling subtest is composed of a dictation task at both levels with word difficulty appropriate to the level given. In addition, Level I Spelling includes copying nonsense figures and name writing. The Reading subtest involves a reading and pronunciation list at both levels, and adds letter reading and recognition at Level I. The Arithmetic portion of the test includes both oral and written tasks at both levels (Lezak, 1976, pp. 232-234). Since the Arithmetic subtest has a very small verbal component, scores from
this portion were not utilized in this study. The mean grade equivalent for Reading and Spelling was used as the verbal standardized performance measure.

The verbal non-standardized performance measure was determined by the Cleaves School Adjustment Scale. These rankings from 0 to 9 are subjectively arrived at by the testing psychologist and are based on the subject's school grades in five academic areas. For the matrix, the mean of the Reading and Spelling rankings was utilized.

The absolute score on the MFD served as the motor memory measure. This test assesses sensory-motor memory by requiring the subject to draw from memory designs that are presented one at a time for five seconds and then removed. Fifteen separate geometric figures of varying complexity designed to detect individuals with brain damage are included in this test. The subject's reproductions are scored for errors according to a point system. The test data yields a raw score which may be interpreted directly or subjected to a correction factor for age and general ability level (Lezak, 1976, pp. 375-376).

The VMI is similar to the MFD in construction and administration, but it does not involve a memory component. It consists of 24 geometric forms of increasing difficulty which are to be copied directly. In this test the figures are not removed from the subject's view. Of these 24 figures, 17 are straight-line, angular configurations while 7 involve
circular elements or discontinuous details. Scoring on each figure is pass/fail and the total score is the number of figures copied successfully prior to three consecutive failures. Like the MFD, the VMI is designed to assess brain damage and identify the learning disabled (Buros, 1978, pp. 1398-1401). Test data yield a raw score and an age equivalent in months and years. The raw score comprised the motor standardized performance measure.

The WISC-R consists of twelve subtests, six of which are considered to be verbal and six non-verbal. The verbal scale subtests are Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span. Each requires the subject to understand a simple statement or question and to formulate a response. Five of the tests involve the giving of an original answer and one (Digit Span) involves the repetition of the examiner's verbalizations. The performance scale subtests are Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, and Mazes. Three of these tests involve the physical assembly of pictures or objects in the correct pattern (Picture Arrangement, Block Design, and Object Assembly), two involve paper-and-pencil tasks (Coding and Mazes), and one (Picture Completion) requires verbal interpretation only. The WISC-R yields Verbal Scale IQ, Performance Scale IQ, and Full Scale IQ scores (Knopf, 1979, pp. 140-141). The motor non-standardized performance measure, the Cleaves Motor Index, is factor analyt-
ically derived from the six separate performance subtest scores according to the following formula:

\[
\frac{[(\text{Picture Completion} - \text{Picture Arrangement}) + 
(\text{Picture Arrangement} - \text{Mazes}) + (\text{Mazes} - \text{Block Design}) + (\text{Block Design} - \text{Object Assembly}) + 
(\text{Object Assembly} - \text{Coding})]}{5}.
\]

**Procedure**

Interpretation of the multitrait-multimethod matrix involves an analysis of the pattern of results appearing therein. A hypothetical 3 X 3 matrix similar to the Campbell and Fiske (1959) example will be used to explain the interpretation of the matrix. The sample matrix is presented in Table 1. Results from the present research were interpreted in an analogous manner.
**TABLE 1**

Hypothetical 3 X 3 Multitrait-Multimethod Matrix Showing Four Conceptual Units

There are four conceptual units represented in the multitrait-multimethod matrix.

1. Reliability diagonals -- The values enclosed in parenthesis are the test-retest reliabilities. The reliabilities are sometimes referred to as the monotrait-monomethod values. In the sample matrix there are three reliability
diagonals, one for each method. In the present research the matrix will not contain reliability diagonals. They are discussed here only for comprehension.

2. Heterotrait-monomethod triangles -- The values enclosed in the solid lines are the intercorrelations between different traits using only one method. Together the reliability diagonals and their adjacent heterotrait-monomethod triangles form the monomethod blocks.

3. Validity diagonals -- The underlined values denote those monotrait-heteromethod values which represent direct measures of convergent validity. There is a validity diagonal for each method and these values are of central importance to validity evaluation.

4. Heterotrait-heteromethod triangles -- The values enclosed in the broken lines are the intercorrelations between different traits using different methods. Together the validity diagonals and their adjacent heterotrait-heteromethod triangles form the heteromethod blocks.

According to Campbell and Fiske, interpretation of a matrix such as this involves an analysis of the following four validity criteria.

1. Values in the validity diagonals should differ significantly from zero and should be large enough to encourage further investigation of the matrix.

2. A validity diagonal value should be higher than the correlation obtained between that variable and any other
variable having neither trait nor method in common. This involves comparing each value in the validity diagonals with selected values directly above, below, and to the sides which are enclosed within the heterotrait-heteromethod triangles. Although common sense dictates that this requirement be met, as it is in Table 1, in the literature it is often overlooked. This criterion was evaluated in the present analysis.

3. A variable should show a higher correlation with an independent measure of that trait than with measures designed to evaluate different traits using the same method. This involves comparing each value in the validity diagonals with those values in the heterotrait-monomethod triangles which represent different traits assessed by the same method. For example, in order to meet this criterion, $r_{A_1 A_2}$ and $r_{A_1 A_3}$ should be greater than $r_{A_1 B_1}$, $r_{B_1 C_1}$, and $r_{A_1 C_1}$. In the sample matrix found in Table 1 variables $A_1$, $B_1$, and $C_1$ meet this requirement totally, while the other variables satisfy it only partially. In individual difference research, it is a rare matrix indeed which can meet this requirement for all variables.

4. The same pattern of trait inter-relationships should appear in all of the heterotrait triangles of both the monomethod and heteromethod blocks. A different pattern indicates possible method bias. The data in the hypothetical matrix shown meet this requirement.

The first of these four criteria is a direct measure of convergent validity. The last three all bear on the concept
of discriminate validity. Considered together, a total pattern leading to the evaluation of construct validity can be obtained.

As was previously stated, a multitrait-multimethod matrix may be of any size, provided that it includes a minimum of two traits measured by a minimum of two methods. The data analysed in this study required a 2 X 3, rather than a 3 X 3 multitrait-multimethod matrix as described in the previous example. This matrix, involving two traits and three methods, is similar in form to a matrix presented by Campbell and Fiske (1959) developed from data reported by Thorndike (1936). This matrix is reproduced in Table 2. Just as in this matrix, the multitrait-multimethod matrix employed in the present study does not include reliability diagonal values. Interpretation of this smaller matrix follows an assessment of the criteria reported above and is included here as another sample of a matrix indicating construct validity.
### TABLE 2
Social Intelligence and Mental Alertness Subtest
Intercorrelations from Thorndike's Data
(N=750)

<table>
<thead>
<tr>
<th></th>
<th>Memory</th>
<th>Comprehension</th>
<th>Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$A_1$</td>
<td>$B_1$</td>
<td>$A_2$</td>
</tr>
<tr>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Intelligence</td>
<td>$A_1$ (        )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental Alertness</td>
<td>$B_1$ .31 (        )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Comprehension</td>
<td></td>
<td>$A_2$ .30 .31 (        )</td>
<td></td>
</tr>
<tr>
<td>Social Intelligence</td>
<td>$B_2$ .29 .38 .48 (        )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mental Alertness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vocabulary</td>
<td></td>
<td>$B_3$ .30 .58 .40 .48 .47 (        )</td>
<td></td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSION

A Pearson Product Moment Correlation Coefficient (Pearson r) was calculated for the scores obtained on each pair of measurement instruments. These correlations appear in the multitrait-multimethod matrix presented in Table 3.

TABLE 3

Verbal and Motor Skills Intercorrelations

(N=25)

<table>
<thead>
<tr>
<th>Memory</th>
<th>Verbal (AVLT) A₁</th>
<th>Motor (MFD) B₁</th>
<th>.28*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stand. Perf.</td>
<td>A₂</td>
<td>B₂</td>
<td>.52*</td>
</tr>
<tr>
<td>Non-Standardized Performance</td>
<td>A₃</td>
<td>B₃</td>
<td>.34a, .31* ,67c ,31</td>
</tr>
</tbody>
</table>

Motor (cleaves Motor Index) B₃-.27 ,01*-.28 ,02 -.24 ( )

a p < .05, b p < .025, c p < .005

* N=15
The pattern of these correlations within the matrix was analyzed according to the validational criteria suggested by Campbell and Fiske.

**Convergent Validity**

Convergent validity is exhibited by a high correlation between maximally independent methods designed to measure the same trait. Validity diagonal values provide a direct measure of convergent validity. Examination of these values revealed three of the six monotrait-heteromethod values to be significantly greater than zero (p < .05). Among assessments of verbal skills, the non-standardized performance measure (Cleaves School Adjustment Scale) showed a high correlation with both the memory measure (AVLT) ($r = .34$), and the standardized performance measure (Verbal WRAT) ($r = .67$). Among assessments of motor skills, the memory measure (MFD) proved to be highly correlated with the standardized performance measure (VMI) ($r = -.56$). The other values in the validity diagonals did not reach acceptable limits of significance.

**Discriminate Validity**

Discriminate validity is exhibited by a low correlation between methods designed to measure different traits. The first of three tests for this involved comparing those validity diagonal values which were earlier found to differ significantly from zero with selected values in the heterotrait-heteromethod triangles. This was a test of the common-sense tenet that a validity value should be higher than the
correlation obtained between that variable and any other having neither trait nor method in common. All three of the validity diagonal values under examination (.34, .67, and -.56) exceeded their corresponding heterotrait-hetero-method comparison values (.34 > .24, -.31, -.27, .31; .67 > .31, .31, -.28, -.52; -.56 > .24, -.52, -.31, .31).

A second assessment of discriminate validity involved comparing validity diagonal values with those values in the heterotrait-monomethod triangles which represent different traits assessed by the same method. This tested the requirement that a variable show a higher correlation with an independent measurement of that trait than with measures designed to evaluate different traits using the same method. As an example of how this was determined, both of the validity diagonal values for the non-standardized performance method of assessing verbal skills (.34 and .67) were required to equal or exceed the value in the heterotrait-monomethod triangle for the non-standardized performance method (-.24). Two of the variables in this study (AVLT and Cleaves School Adjustment Scale) met this requirement totally, two met it partially (Verbal WRAT and MFD) and two (VMI and Cleaves Motor Index) did not meet it at all.

The third evaluation of discriminate validity involved a test for method bias. If the values in the monomethod triangles substantially exceed those in the heteromethod triangles, this indicates the possibility that systematic factors
are responsible for score variations. In the multitrait-multimethod matrix presented here, the values in the hetero-trait-monomethod triangles excluded this possibility in two out of three cases. Only the standardized performance method, with a monomethod triangle value of .60, was indicative of possible method bias.
GENERAL DISCUSSION

The results of this investigation provided evidence for convergent validity. Three of the six monotrait-heteromethod correlations were found to be reliable. These included the non-standardized performance measure (Cleaves School Adjustment Scale) correlated with the memory measure (AVLT) and the standardized performance measure (Verbal WRAT). For verbal skills the pattern of correlations between these three independent measures provided evidence for convergent validity. In addition, for motor skills, the memory measure (MFD) was significantly related to the standardized performance measure (VMI). This result was not unexpected because these two measures share a similar construction and administration.

The evidence for discriminate validity was developed employing three sets of comparisons. First, we concentrated on analyzing the three significant correlations in the validity diagonals. In each case, the value in the validity diagonal exceeded its corresponding heterotrait-heteromethod comparison values. As far as these significant validities are concerned, this pattern of results constitutes evidence for discriminate validity.

A second comparison involved the validity diagonal values and their corresponding comparison values in the
heterotrait-monomethod triangles. Two of the measures, the AVLT and the Cleaves School Adjustment Scale, provided evidence for discriminate validity.

Finally, the possibility of method bias was assessed. The pattern of correlations indicated an absence of systematic method bias for two of the three methods. However, for the standardized performance method, the possibility of method bias exists.

In summary, the analysis provided considerable evidence for both convergent and discriminate validity. The pattern of results, however, is not unequivocal. For example, only half of the validity diagonal values were statistically significant. Evidence for convergent validity for verbal skills was not found when measured with memory and standardized performance methods (AVLT and Verbal WRAT) together. Furthermore, for motor skills, correlations between the non-standardized performance measures failed to indicate convergent validity.

Keeping in mind the low sample size, the results encourage the use of the VMI and the MFD to measure motor skills, and the Cleaves School Adjustment Scale along with either the Verbal WRAT or the AVLT for measuring verbal skills. A comparable vote of confidence is not justified for measuring verbal skills using both the Verbal WRAT and the AVLT or for measuring motor skills using the Cleaves Motor Index.
In light of the fact that the source of subjects continues to be available, the favorable pattern of results in this analysis clearly indicates the fruitfulness of pursuing this study utilizing a larger sample size.
REFERENCES


