DOCUMENT RESUME

ED 311 059 TM 013 873

AUTHOR Byrne, Barbara M.
TITLE Four Multitrait-Multimethod Approaches to Examining the Construct Validity of Adolescent Self-Concept within and across Gender.
PUB DATE Aug 89
NOTE 36p.; Paper presented at the Annual Meeting of the American Psychological Association (97th, New Orleans, LA, August 1989).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Adolescents; Analysis of Variance; Comparative Analysis; *Construct Validity; Estimation (Mathematics); Evaluation Methods; Factor Analysis; Foreign Countries; Grade 11; Grade 12; High Schools; *High School Students; *Multitrait Multimethod Techniques; *Self Concept Measures; *Sex Differences; Test Validity
IDENTIFIERS Campbell Fiske Matrix; Composite Direct Product Model (Browne); Confirmatory Factor Analysis; Ontario (Ottawa)

ABSTRACT

The construct validity of a multidimensional adolescent self-concept was estimated separately for males and females, and the results were compared across gender. The application of four multitrait-multimethod (MTMM) approaches to estimating construct validity was demonstrated, and the scope and consistency of the findings derived from each were compared. Methodological approaches included: (1) the Campbell-Fiske model of D. T. Campbell and D. W. Fiske (1959); (2) the analysis of variance model of M. J. Kavanagh et al. (1971); (3) the confirmatory factor analysis model of K. G. Joreskog (1969); and (4) the composite direct product model of M. W. Browne (1984). These procedures were applied to data comprising four self-concept traits (general, academic, English, and mathematics) as measured by three measurement scales (Likert, semantic differential, and Guttman). The sample consisted of 412 males and 420 females in grades 11 and 12 from two suburban high schools in Ottawa (Ontario). Results derived from the composite direct product model are less than optimal. Other results demonstrate that adolescent males and females can differ with respect to various aspects of construct validity related to multidimensional facets of self-concept. Seven tables present study data. (SLD)

Four Multitrait-multimethod Approaches to Examining the Construct Validity of Adolescent Self-concept Within and Across Gender

Barbara M. Byrne
University of Ottawa

Paper presented at the American Psychological Assocation Annual Meeting, New Orleans, 1989

The author wishes to thank Michael W. Browne for making the MUTMUM computer program available, and Richard D. Goffin for his invaluable assistance in the use of this program and in the interpretation of particular results based on the the Composite Direct Product Model.

Abstract

The purposes of the study were (a) to estimate the construct validity of a multidimensional adolescent self-concept separately for males and females and to compare the results across gender, and (b) to demonstrate the application of four MTMM approaches to estimating construct validity and to compare the scope and consistency of the findings derived from each. Methodological approaches included the Campbell-Fiske model (1959), the ANOVA model (Kavanagh, MacKinney, & Wolens, 1971), the confirmatory factor analysis model (Joreskog, 1969), and the composite direct product model (Browne, 1984). These procedures were applied to data comprising four self-concept traits (general, academic, English, mathematics) as measured by three measurement scales (Likert, semantic differential, Guttman) for 832 grades 11 and 12 high school students (males = 412; females = 420).

Four Multitrait-multimethod Approaches to Examining the Construct Validity of Adolescent Self-concept Within and Across Gender

The purposes of the present study were twofold. First, to examine the construct validity of adolescent self-concept (SC) within gender and to compare the results across gender. Second, to demonstrate the application of four multitrait- multimethod (MMTM) approaches to determining construct validity and to compare the scope and consistency of the findings derived from each.

Of late, an increasing number of substantive studies have investigated the possibility of gender differences in adolescent SC. In particular, interest has focused on general SC, academic SC, and the more specific subject-matter SCs related to mathematics and English. An important assumption in testing for mean differences in SC, however, is the equivalency across groups of (a) the structure of the underlying construct, and (b) the validity and reliability of the measures.

Substantive research findings bearing on gender differences in adolescent SC are, in general, inconsistent (for reviews see Byrne, 1984; West, Fish & Stevens, 1980; Wylie, 1974, 1979). Although findings related to the more specific facets of English SC and mathematics SC are somewhat more consistent in demonstrating higher English SCs for girls, albeit higher mathematics SCs for boys (Byrne & Shavelson. 1986; Marsh, Parker & Barnes, 1985; Meece, Parson, Kaczala, Goff &

Fetterman, 1982), there is still some uncertainty concerning the equivalence of relations among these facets of the construct. Byrne and Shavelson (1987), in explaining the discrepant SC findings, cited three major limitations of earlier studies: (a) the lack of a clear theoretical rationale, (b) the use of psychometrically limited instrumentation, and (c) the use of simplistic or inappropriate methodological procedures.

The inconsistencies in SC research, however, can also be explained by gender differences in the construct validity of both the multidimensional SC facets, and the instruments designed to measure them. Indeed, implicit in multigroup comparisons is the assumption of group-invariant construct validity; nonetheless, such equivalence is rarely tested directly. Although Byrne and Shavelson (1987) investigated invariance of the measurement and structure of adolescent SC, they did not examine the equivalence of convergent and discriminant validities nor method effects associated with the scale format of particular measuring instruments (see Fiske, 1987; Spector, 1987).

Campbell and Fiske (1959) posited that claims of construct validity must be accompanied by evidence of both convergent and discriminant validity (see also, Messick, 1981). As such, a measure should correlate highly with other measures to which it is theoretically linked (convergent validity), and correlate negligibly with those that are theoretically unrelated (discriminant validity). Method effects represent bias that can derive from use of the same measurement method in the assessment of different constructs. Such bias impinges on the magnitude of trait relations, thus rendering results spurious. Method bias has been shown to be differentially influenced by gender (Hamilton, 1981). One procedure for determining evidence of convergent and discriminant validity, and for detecting method bias is to examine data within a MTMM matrix framework.

The seminal work of Campbell and Fiske (1959) provided the initial means to examining evidence of construct validity. These researchers proposed that measures of multiple traits be assessed by multiple methods and that all trait-method correlations be arranged in a MTMM matrix. The assessment of construct validity then focuses on comparisons among three blocks of correlations: (a) scores on the same traits measured by different methods (monotrait-heteromethod values i.e., convergent validity), (b) scores on different traits measured by the same method (heterotrait-monomethod values i.e., discriminant validity) and, (c) scores on different traits measured by different methods (heterotrait-heteromethod values i.e., discriminant validity). Specific criteria guide the inspection of these values.

Over the past decade, researchers have noted several shortcomings in the Campbell-Fiske procedure and have proposed alternative approaches to the analysis of MTMM matrices. One such methodology summarizes the MTMM data within an analysis of variance (ANOVA) three-way (subjects x traits x methods) factorial design (Kavanagh, MacKinney, & Wolens, 1971). As such, four major effects are of interest, and relate to variance due to: (a) subjects (convergent validity), (b) subjects by traits (discriminant validity), (c) subjects by methods (method bias), and (d) error; all effects are tested statistically.

A third approach to determining construct validity is the analysis of covariance structures using the confirmatory factor analytic (CFA) methodology proposed by Joreskog (1969). This procedure is purported to have several advantages over the Campbell-Fiske approach. First, the MTMM matrix is explained in terms of the underlying latent constructs, rather than the observed variables. Second, the evaluation of convergent and discriminant validities can be made at both the matrix and individual parameter levels. Third, hypotheses related to convergent and discriminant validity can be tested statistically, based on a series of hierarchically nested models. Finally, separate estimates of variance due to traits, methods, and error/uniquenesses are provided.

A fourth approach to determining construct validity derives from the decomposition of MTMM matrices using the Composite Direct Product (CDP) model proposed by Browne (1984). This methodology, as with the CFA model, is based on the anlaysis of covariance structures. Unlike the CFA model, however, the CDP model is based on a component analytic, rather then on a factor analytic methodological approach. This procedure comprises a more parsimonious parameterization of general trait and method variance than the CFA model and, as such, is less subject to problems of indeterminacy (Wothke, 1987).

The present study extends the work of Byrne and Shavelson (1987) by reexamining the same data within a MTMM framework using four conceptually different approaches to the data analyses; each procedure providing more information and being a more powerful test of hypotheses than its predecessor. Specifically, the study examines the extent to which the convergent and discriminant validity of four SC traits (general SC, academic SC, English SC, mathematics SC), and the method effects associated with three measurement scales (Likert, semantic differential, Guttman) are equivalent across gender. MTMM analyses will include evaluations based on (a) Campbell-Fiske criteria, (b) ANOVA model (Kavanagh, MacKinney & Wolins, 1971), (c) CFA model (Joreskog, 1969), and (d) CDP model (Browne, 1984).

Method

Sample and Procedure

Subjects were 832 (412 males, 420 females) grades 11 and 12 students from two suburban high schools in Ottawa, Ontario. A battery of SC instruments were administered to intact classroom groups during one 50-minute period. Testing was completed approximately two weeks following the April report cards. Students therefore had the opportunity of being fully cognizant of their academic performance prior to completing the tests for the study; this factor was considered important in measuring academic and subject-matter SCs.

Instrumentation

The SC battery consisted of 12 measures; three different measurement scales (Likert, semantic differential, Guttman) for each of general SC, academic SC, English SC, and mathematics SC. All were self-report instruments designed specifically for use with a high school population. The Likert scales comprised subscales of the Self Description Questionnaire III (SDQ III; Marsh & O'Neill, 1984) designed to measure general SC, academic SC, English SC and mathematics SC; the semantic differential measures included subscales of the Affective Perception Inventory (API; Soares & Soares, 1979) designed to measured these same SC facets. Finally, the Guttman scales were the Self-esteem Scale (SES; Rosenberg, 1965) designed to measure general SC, and the Self-concept of Ability Scales (SCAS; Brookover, 1962) adapted for separate measurements of academic SC, and the subject-matter SCs.

Analysis of the Data

The data were analyzed separately for each MTMM model. First, zero-order correlations among all measurements were computed and arranged in a separate MTMM matrix for males and for females; Campbell-Fiske criteria were used to assess evidence of convergent and discriminant validity both within and across gender. Second, trait-method correlations were decomposed into variance components and summarized within an ANOVA framework; variance explaining trait and method factors was compared across gender. Finally, using LISREL covariance factor (CFA model) and covariance component (CDP model) analytic procedures, a 7-factor (4 traits, 3 methods) model of the data was proposed and tested, separately, for males and females; estimates of variance due to each trait and method were compared across gender.

Results

Campbell-Fiske Criteria

Campbell and Fiske (1959) proposed four criteria for evaluating convergent and discriminant validity. These criteria are:

1. The convergent validities should be significantly different from zero and sufficiently large to warrant further investigation of validity.

2. The convergent validities should be higher than correlations between different traits assessed by different methods (heterotrait-heteromethod blocks).

3. The convergent validities should be higher than correlations between different traits assessed by the same method (heterotrait-monomethod blocks).

4. The pattern of correlations between different traits should be the same in both the heteromethod and monomethod blocks.

For each sex, comparisons of various blocks of correlations involved determining the proportion of times that these criteria were satisfied; these matrices of zero-order correlations are presented in Table 1, together with the means, standard deviations, and internal consistency alpha reliabilities. Results are entered below the main diagonal for the males, and above the main diagonal for females.

Insert Table 1 about here

Criterion 1. Convergent validities were all statistically significant (p <.05; mean r=.69) for both males and females.

Criterion 2. Convergent validities were consistently higher than correlations between different traits assessed by different methods (heterotrait-heteromethod triangles) for both males (34 of 36 comparisons) and females (35 of 36 comparisons).

Criterion 3. Although convergent validities, for the most part, were consistently higher than correlations between different traits measured by the same method (heterotrait-monomethod triangles) for females (13 of 18 comparisons), this was not so males; only 10 of the 18 comparisons met this criterion. In particular, the semantic differential and Guttman scales exhibited some method bias; whereas the former was marginally stronger for males, the reverse was true for females.

Criterion 4. Within each sex, the pattern of correlations among the different traits was fairly similar across methods. Across gener, however, ASC differed in its relation with ESC and MSC; whereas ASC correlated higher with MSC than with ESC for males, the latter relation was almost equivalent for females.

Analysis of Variance

As an alternative to the Campbell-Fiske approach to analyzing MTMM data, Kavanagh et al. (1971) proposed an incomplete 3-way analysis of variance (ANOVA) procedure that decomposes the matrix elements into four sources of variance: (a) subjects variance, the overall agreement of subjects over all traits (i.e., convergent validity), (b) subjects by traits variance, the extent to which subjects are differentiated by various trait measures (i.e., discriminant validity related to traits), (c) subjects by methods, the extent to which method effects impinge upon the methods used (i.e., discriminant validity related to methods), and (d) error variance, the subjects by traits by methods interaction. As such, subjects are considered as random, while traits and methods are considered as fixed. Furthermore, an index can be computed for determining the relative contribution of each variance component to the total variation attributable to convergent and discriminant validity, method effects, and error; the index can be standarized for purposes of MTMM comparisons across groups. The ANOVA results are presented in Table 2

Insert Table 2 about here

In reviewing these findings, we see an interesting difference between adolescent males and females with respect to the main effect for subjects. and the subjects by traits interaction, respectively. Although both were moderately strong for both sexes, subjects accounted for slightly more variance for males (.43) than for females (.37), while the reverse was true for the subjects by traits interaction (females = .45; males = .37). These results indicate that while adolescent males were better discriminated by the trait-method combinations employed than were females, discrimination between SC facets was slightly higher for females, than for males. Nonetheless, for both sexes, these two components accounted for the largest portion of variance within the analyses. Interestingly, findings bearing on the subjects by methods interaction for both sexes indicated negligible effects due to method effects (males = .04; females = .05); these results suggest no operation of method bias within any of the three methods employed, for either sex.

Although the above findings suggested some differences across gender, direct comparison must derive from standardized measures since variance components are influenced by within gender error. An examination of these standardized indices, however, substantiated the findings related to evidence of convergent and discriminant validity, albeit a relatively weak method effect appeared to operate for both sexes.

Confirmatory Factor Analysis

For each sex, the 7-factor model was tested for convergent and discriminant validity by means of (a) comparisons with alternatively specified models, and (b) examination of individual parameter estimates. All CFA analyses were conducted using LISREL VI (Joreskog & Sorbom, 1985).

In covariance structure analysis, the extent to which a proposed model fits the observed data is based on the likelihood ratio test. However, problems related to the dependency of χ^2 on sample size have led to recommendations that the assessment of model fit be based on multiple criteria that include a measure of practical fit (Bentler & Bonett, 1980; Joreskog, 1971; Joreskog & Sorbom, 1985; Marsh & Hocevar, 1985; Widaman, 1985). This was accomplished in the present study by using (a) the χ^2 likelihood ratio, (b) the χ^2/degrees of freedom (df) ratio, (c) the Bentler & Bonett normed index (BBI), (d) T-values and modification indices provided by the LISREL VI program, and (e) knowledge of substantive and theoretical research in this area.

To establish various validity criteria, the proposed 7-factor model was tested against a series of more restrictive models in which specific parameters were either eliminated or constrained to equal zero. Since the difference in χ^2 ($\Delta\chi^2$) is itself χ^2-distributed, with degrees of freedom equal to the difference in degrees of freedom for the two models, the fit differential between comparison models can be tested statistically. A significant $\Delta\chi^2$ argues for the superior statistical fit of the less restrictive model in the sense that it exhibits a lower non-centrality parameter (see Saris & Stronkhorst, 1984). However, despite widespread acceptance and practise of this approach to model fitting, researchers are urged to exercise caution in implying certainty where none exists; unless one of the models fits perfectly, any interpretation of a non-zero difference remains problematic. (For a more detailed discussion of MTMM model comparisons, see Widaman, 1985, and of model application, see Byrne, 1989; Marsh & Hocevar, 1983). In addition to model comparisons at the matrix level, individual parameter estimates for trait and method factor loadings, trait intercorrelations, method intercorrelations, and estimated error/uniquenesses were examined with respect to magnitude and statistical significance.

Examination of matrices. Goodness-of-fit indices for the series of MTMM models tested for evidence of convergent and discriminant validity are presented in Table 3. Model 1 is the most restrictive model, representing the null hypothesis that each observed measure is an independent factor; it serves as the null model against which competing models are compared in order to determine the BBI. Models 2-4 represent decreasingly restrictive models, such that Model 4 is the least restrictive, having both correlated traits and correlated methods. Model 4 is termed the baseline model since it represents hypothesized relations among the traits and methods; typically, it demonstrates the best fit to the data.

Insert Table 3 about here

For both males and females, as expected, Model 4 represented the best fit to the data. Due to known problems of estimation (see Widaman, 1985), additional fitting of this hypothesized model was not conducted; as such, Model 4 included traits and methods, each correlated only among themselves. (Due to technical difficulties with the LISREL program, trait-method correlations cannot be tested (Widaman, 1985)).

As noted earlier, one criterion related to evidence of construct validity is that convergent validities should be substantial and statistically significant. Thus, if we compare a model in which traits are specified, with one in which they are not, a significant difference in χ^2 between the two models argues for evidence of convergent validity. As such, Model 4, in which traits were freely estimated, was compared with Model 5 in which no trait factors were specified. As shown in Table 4, the $\Delta\chi^2$ was highly significant for both sexes, albeit females demonstrated somewhat stronger evidence of convergent validity than males.

Insert Table 4 about here

In a similar manner, we can test for evidence of discriminant validity. For example, complete discriminant validity of traits argues for their zero intercorrelations. Thus, by comparing a model in which traits are allowed to correlate freely, with one in which trait correlations are constrained to 1.00 (i.e., perfectly correlated), we can test for discriminant validity. As such, Model 4, in which traits were allowed to correlate freely, was compared with Model 6 which specified perfect correlations among traits. The highly significant $\Delta\chi^2$, as shown in Table 2, again indicated strong evidence of trait discriminant validity for both sexes.

Evidence of method effects (i.e., lack of discriminant validity among the methods) can be tested by comparing a model in which method factors are freely estimated, with one in which no method factors are specified; a significant $\Delta\chi^2$ argues for the lack of discriminant validity and, thus, for evidence of method bias. Accordingly, Model 4 was compared with Model 2 in which no method factors were specified. Although this comparison yielded statistically significant $\Delta\chi^2$'s for both sexes, the differential values were of much lesser magnitude than for the traits. Nonetheless, method discriminability was evident for both males and females.

Finally, to determine the extent to which each measurement scale was contributing to method effects, Model 4 was further compared with three additional models, each of which eliminated one of the three methods. Interestingly, for both males and females, the Likert scales appeared to contribute least to method bias, whereas the Guttman scales contributed the most. Overall, method effects appeared comparable across gender (c.f. Hamilton, 1981).

Examination of parameters. A more precise assessment of trait- and method-related variance can be ascertained by examining the individual parameter estimates as specified for the hypothesized model (Model 4). Of specific interest are the magnitude of the factor loadings and the factor correlations. These results are presented in Tables 5 and 6 for males and females, respectively. The magnitude of the trait loadings are consistent with earlier convergent validity findings determined at the matrix level (see Table 4); all loadings were statistically significant for both sexes.

Insert Tables 5 and 6 about here

Method factor loadings appear relatively similar across gender, with two minor exceptions. First, method-related variance was slightly higher for males than for females, with respect to the Likert measure of general SC (.58 vs. .42) Second, method loadings related to academic SC consistently differed across gender; whereas method variance associated with the Likert and semantic differential measures were somewhat higher for males (.58, .41, respectively) than for females (.42, .31, respectively), the reverse was true for the Guttman measure (males .53; females .62).

Discriminant validity of traits and methods can be evidenced by examining the factor correlation matrices. Results generally supported earlier findings reported at the matrix level (see Table 2). As such, discriminant validity related to the traits tended to be slightly stronger for females (mean r=.41), than for males (mean r=.50). Nonetheless, all values were more than three standard errors from 1.00, indicating evidence of discriminant validity. Indeed, Marsh and Hocevar (1983) noted that only when factor correlations approach unity should researchers be concerned about a lack of discriminant validity with respect to attitudinal constructs such as SC.

Widaman (1985) has noted that, despite Campbell and Fiske's admonition for the use of maximally dissimilar methods where multiple measures are employed, little attention has been paid to the discriminability of method effects. With one exception, method correlations (see Tables 5 and 6) indicated only modest method discriminability for both males and females. The correlation of .87 between the Likert and Guttman scales for females was less than 1 standard error from 1.00, indicating a strong method effect related to the combination of these two methods; all other correlations were more than two standard errors from 1.00, supporting evidence of discriminant validity.

Composite Direct Product Analyses

As noted above, the estimation of trait-method correlations for the CFA model is problematic and not recommended (see Widaman, 1985). Indeed, Campbell and O'Connell (1967, 1982) have demonstrated that trait and method factors may interact in a multiplicative, rather than an additive fashion such that trait correlations become inflated when both share the same method. In other words, the stronger the relation between two traits, the more their correlation is likely to increase when the method is shared; conversely, independent or near-independent traits remain so, despite an identical method of measurement. Of importance, then, is that method effects tend to exaggerate correlations between traits that are strongly intercorrelated.

In a seminal paper that addressed this issue, Browne (1984) presented a modified CFA approach designed to accommodate the trait-method interaction effect described by Campbell and O'Connell (1967, 1982). As such, Browne, based on earlier research by Swain (cited in Browne, 1984), proposed the CDP model. The major tenet of this approach to MTMM analyses is that the true correlation matrix of observed data is the product of trait and method factors. Browne argued that method effects act to inflate or deflate relations between particular traits.

In order to test these hypotheses, Browne (1984) developed MUTMUM, a computer program that assumes an underlying CDP structure of observed MTMM data. Results based on the present data were analyzed using this program, and are presented in Table 7. The component correlation matrices $\hat{P}_T$ and $\hat{P}_M$ represent correlations among the four traits and 3 methods, respectively. Since the elements in these matrices represent multiplicative component correlations corrected for attenuation, they are expected to be higher than correlations between observed scores as presented in Table 1 (Browne, 1984); correlations of 1.00, notwithstanding, are very disturbing. These values are indicative of boundary parameters, which are automatically fixed to 1.00 by the MUTMUM program (Goffin, personal communication, August, 1989).

Insert Table 7 about here

Interpretation of results from the CDP model analyses focuses on the trait and method effects as reflected in their respective correlation matrices. To assess these effects, comparison is made with an extreme model in which all traits or all methods are perfectly correlated; the former argues for no trait effects (i.e., no trait variance), while the latter argues for no method effects (i.e., no method variance). Thus, if the mean correlation in $\hat{P}_T$ is less than the mean correlation in $\hat{P}_M$ the claim can be made that, as a group, trait factors are impacting to a larger degree than are method factors on the MTMM matrix elements (Goffin, 1987); the reverse, of course, substantiates the greater impact of method factors. Since the mean trait correlation for both males (mean r = .30) and females (mean r = .57) was less than the mean method correlation (males, mean r = . 36; females, mean r = .81), this finding argues for the superiority of trait effects over method effects. Nonetheless, while method effects appear to be fairly weak for both sexes, the impact of trait effects seems to be much stronger for females.

Cudeck (1988) has recently shown that results from CDP model analyses can be evaluated in a straightforward fashion in terms of the Campbell-Fiske criteria. (Since these criteria have been described earlier in the paper, they are not repeated here). As such, Criterion 1 may be considered satisfied if all elements in $\hat{P}_M$ are large (i.e., convergent validities are large when methods are highly correlated). With one exception related to females (r = .44), this criterion seems otherwise adequately met by both sexes.

Criterion 2 can be evaluated by examining the $\hat{P}_T$ matrix.

Cudeck (1988) argues that this criterion is satisfied if (a) there are no trait correlations >1.00, and (b) the model seems appropriate to the data. Given the finding of two correlations equal to 1.00 for females, and a totally unacceptable fit to the data for both males (χ^2 (45) = 1319.46) and females (χ^2 (45) = 1415.50), the present data failed to meet this criterion thus providing insufficient evidence of discriminant validity related to SC traits.

The satisfaction of Criterion 3 argues for the discriminant validity of method factors, and hence, the unimportance of method effects. Cudeck suggests that if the elements in the $\hat{P}_M$ matrix exceed those in the $\hat{P}_T$ matrix, this criterion can be considered met; indeed, it appears to have been satisfied for both males and females.

Finally, claims that Criterion 4 has been satisfied are evidenced by a well-fitting model which indicates an accurate description of the data (Cudeck, 1988). Since the fit of the CDP model to both the male and female data, as noted above, is completely unsatisfactory, we therefore conclude that Criterion 4 was not met. Indeed, Browne (1984) has noted that, given the more parsimonious parameterization of the CDP model compared with the restricted CFA model, the former is more easily rejected by goodness-of-fit tests. Nonetheless, in light of the extremely poor model fit for both sexes, the author recommends that no substantive interpretations be made from the present CDP results; they are therefore included here for purposes of illustration only.

Conclusions

Four approaches to the analysis of MTMM data comprising four traits and three methods, have been described and illustrated using SC data from high school adolescent males and females. Given the elaboration of contingencies associated with each methodological procedure, it is hoped that researchers will benefit substantially from the applications used here. Indeed, by demonstrating the range of information available, the four MTMM approaches presented here should aid researchers in (a) assessing the strengths and weaknesses associated with each procedure, and (b) selecting the methodological approach most appropriate to their own research. Although results derived from the CDP model were less than optimal and thus require further investigation, the illustration of this procedure was considered important since the literature is presently void of its application to real (versus simulated) data.

Substantively, the paper has demonstrated that adolescent males and females can differ with respect to various aspects of construct validity related to multidimensional facets of SC. Admittedly, certain findings appeared to be method-specific (e.g., discriminant validity related to traits and methods); further investigation of these results remain the task of future research that is currently underway. Nonetheless, one consistent gender difference bears importantly on the structure of the construct itself - the relation between ESC and MSC with ASC; whereas for females, ESC is more closely related to ASC than is MSC, the reverse is true for males. Indeed, these findings have been repeatedly supported by a number of independent studies (for a review, see Marsh, in press). Such gender difference findings should make researchers and practitioners more cognizant of potential psychometric differences in the measurement of adolescent SC across gender; such differences bear importantly both on the selection of assessment instruments and on interpretative meanings assigned to their rendered scores.

References

Bentler, P.M. & Bonett, D.G. (1980). Significance tests and goodness-of-fit in the analysis of covariance structures. Psychological Bulletin, 88, 588-606.

Brookover, W.B. (1962). Self-concept of Ability Scale. East Lansing, MI: Educational Publication Services.

Browne, M.W. (1984). The decomposition of multitrait-multimethod matrices. The British Journal of Mathematical and Statistical Psychology, 37, 1-21.

Byrne, B.M. (1984). The general/academic self-concept nomological network: A review of construct validation research. Review of Educational Research, 54, 427-456.

Byrne, B.M. (1989). A Primer of LISREL: Basic applications and programming for confirmatory factor analytic models. New York: Springer-Verlag.

Byrne, B.M. & Shavelson, R.J. (1986, April). On gender differences in the structure of adolescent self-concept. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Byrne, B.M. & Shavelson, R.J. (1987). Adolescent self-concept: Testing the assumption of equivalent structure across gender. American Educational Research Journal, 24, 365-385.

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

Campbell, D.T. & O'Connell, E.J. (1967). Methods factors in multitrait-multimethod matrices: Multiplicative rather than additive? Multivariate Behavioral Research, 2, 409-426.

Campbell, D.T. & O'Connell, E.J. (1982). Methods as diluting trait relationships rather than adding irrelevant systematic variance. In D. Brinberg & L.H. Kidder (Eds.), New directions for methodology of social and behavioral science: Forms of validity in research (pp. 93-111). San Francisco: Jossey-Bass.

Cudeck, R. (1988). Multiplicative models and MTMM matrices. Journal of Educational Statistics, 13, 131-147.

Fiske, D.W. (1987). Construct invalidity comes from method effects. Educational and Psychological Measurement, 47, 285-307.

Goffin, R.D. (1987). The analysis of multitrait-multimethod matrices: An empirical and Monte Carlo comparison of two procedures. Unpublished doctoral dissertation, University of Western Ontario, London, Canada.

Hamilton, R.C. (1981). Sex differences in self-report errors: A Note of Caution. Journal of Educational Measurement, 18, 221-228.

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

Joreskog, K.G. & Sorbom, D. (1985). LISREL VI: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood. Mooresville, IN: Scientific Software.

Kavanagh, M.J., MacKinney, A.C., & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analyses of ratings. Psychological Bulletin, 75, 34-49.

Marsh, H.W. (in press). A multidimensional hierarchical model of self-concept: Theoretical and empirical justification. Educational Psycholcgy Review.

Marsh, H.W. & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. Journal of Educational Measurement, 20, 231-248.

Marsh, H.W. & O'Neill, R. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. Psychological Bulletin, 97, 562-582.

Marsh, H.W. & O'Neill, R. (1984). Self Description Questionnaire III: The construct validity of multidimensional self-concept ratings by late adolescents. Journal of Educational Measurement, 21, 153-174.

Marsh, H.W., Parker, J., & Barnes, J. (1985). Multidimensional adolescent self-concepts: Their relationship to age, sex, and academic measures. American Educational Research Journal, 22, 422-444.

Meece, J.L., Parsons, J.E., Kaczala, C.M., Goff, S.B., & Futterman, R. (1982). Sex differences in math achievement: Toward a model of academic choice. Psychological Bulletin, 91, 324-348.

Messick, S. (1981). Constructs and their vicissitudes in educational and psychological measurement. Psychological Bulletin, 89, 575-588.

Rosenberg, M. (1965). Society and the adolescent self-image. Princeton: Princeton University Press.

Saris, W. & Stronkhorst, H. (1984). Causal Modelling in nonexperimental research. Amsterdam: Sociometric Research Foundation.

Soares A.T. & Soares, L.M. (1979). The Affective Perception Inventory -Advanced Level. Trumbell, CT: ALSO.

Spector, P.E. (1987). Method variance as an artifact in self-reported affect and perceptions at work: Myth or significant problem? Journal of Applied Psychology, 72, 438-443.

West, C.K., Fish, J.A., & Stevens, R.J. (1980). General self-concept, self-concept of academic ability and school achievement: Implications for "causes" of self-concept. The Australian Journal of Education, 24 (2), 194-213.

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

Wothke, W. (1987, April). Multivariate linear models of the multitrait-multimethod matrix. Paper presented at the American Educational Research Association Annual Meeting, Washington.

Wylie, R. (1974). The self-concept: A Review of methodological considerations and measuring instruments. Lincoln: University of Nebraska Press.

Wylie, R. (1979). The self-concept Volume 2: Theory and research on selected topics. Lincoln: University of Nebraska Press.

Table 1

Multitrait-multimethod Matrix of Zero-order Correlations Among Self-Concept Measures for Males and Females[a]

	Likert (SDQ III)				Semantic Differential (API)				Guttman (SES)	(SCAS)		
Measure	GSC	ASC	ESC	MSC	GSC	ASC	ESC	MSC	GSC	ASC	ESC	MSC
Likert												
GSC	--	.29	.28	.11	.66	.52	.15	.12	.83	.19	.17	.08
ASC	.35	--	.38	.38	.30	.57	.40	.41	.33	.66	.47	.36
ESC	.29	.38	--	.07	.21	.33	.68	.03	.31	.34	.55	.04
MSC	.25	.58	.11	--	.14	.33	.04	.86	.16	.38	.03	.82
Semantic Differential												
GSC	.64	.33	.25	.25	--	.60	.19	.19	.68	.14	.14	.08
ASC	.54	.64	.38	.45	.66	--	.40	.39	.57	.50	.35	.32
ESC	.17	.43	.72	.14	.27	.44	--	.07	.21	.41	.66	.03
MSC	.25	.60	.20	.88	.29	.53	.29	--	.18	.41	.05	.81
Guttman												
GSC	.77	.34	.30	.28	.60	.57	.22	.31	--	.26	.23	.12
ASC	.25	.70	.34	.59	.31	.59	.36	.62	.31	--	.63	.51
ESC	.10	.41	.51	.14	.20	.37	.63	.20	.16	.55	--	.08
MSC	.21	.53	.13	.83	.25	.47	.16	.81	[illegible]	.68	.23	--

Table 1 (cont'd)

Multitrait-multimethod Matrix of Zero-order Correlations Among Self-Concept Measures for Males and Females[a]

	Likert (SDQ III)				Semantic Differential (API)				Guttman (SES)		(SCAS)	
Measure	GSC	ASC	ESC	MSC	GSC	ASC	ESC	MSC	GSC	ASC	ESC	MSC
Males (n = 412)												
M	76.41	52.98	55.62	49.22	78.22	71.17	57.92	47.40	32.02	28.36	26.82	26.24
SD	13.86	13.38	10.02	15.98	9.44	10.00	11.09	11.76	4.89	5.94	5.93	7.87
α	.94	.89	.78	.92	.86	.85	.88	.95	.85	.88	.90	.94
Females (n = 420)												
M	75.35	57.62	57.93	44.42	75.34	74.15	63.17	43.77	30.73	28.89	28.87	24.24
SD	14.52	11.01	9.47	16.34	8.85	8.65	10.70	11.07	5.02	4.87	5.32	7.09
α	.94	.88	.79	.93	.86	.83	.88	.94	.88	.85	.89	.94

[a] Correlations for males are below the main diagonal, and for females, above the main diagonal. All correlations > .07 are significant ($p < .05$).

Note: The underlined values are convergent validities. The values in solid triangles are discriminant validities (heterotrait-monomethod correlations); those in broken triangles are discriminant validities (heterotrait-heteromethod correlations).

α = alpha reliability coefficient; GSC = general self-concept (SC); ASC = academic SC; ESC = English SC; MSC = mathematics SC; SDQ III = Self Description Questionnaire III; API = Affective Perception Inventory; SCAS = SC of Ability Scale.

Table 2

Analysis of Variance Decomposition of the Multitrait-multimethod Matrix for Males and for Females

Source of Variance	SS	df	MS	F	Variance Component	Standardized Comparison Index
			Males (n = 412)			
Subjects (A) (Convergence)	2234.69	411	5.44	20.92***	.43	.62
A x Traits (B) (discrimination)	1700.74	1233	1.38	5.31	.37	.59
A x Methods (C) (method/halo)	360.91	822	.44	1.69	.04	.13
A x B x C	647.66	2466	.26	--	.26	--
			Females (n = 420)			
Subjects (A) (convergence)	1975.68	419	4.72	16.74	.37	.58
A x Traits (B) (discrimination)	2021.04	1257	1.61	5.96	.45	.63
A x Methods (C) (method/halo)	372.96	838	.45	1.65	.05	.16
A x B x C (error)	670.32	2514	.27	--	.27	--

Note: All F-values are significant at $p < .001$.

Table 3

Goodness-of-fit Indices for Multitrait-multimethod Models

		Males (n = 412)			Females (n = 420)		
	Model	χ^2	df	BBI	χ^2	df	BBI
1	12 uncorrelated factors (null model)	3740.55	66	—	3598.79	66	—
2	4 correlated trait factors no method factors	474.43	48	.873	430.44	48	.880
3	4 correlated trait factors 3 uncorrelated method factors	237.04	36	.937	199.03	37	.945
4	4 correlated trait factors 3 correlated method factors (baseline model)	182.70	34	.951	126.37	34	.965
5	no trait factors 3 correlated method factors	1867.18	52	.501	2251.15	52	.374
6	4 perfectly correlated trait factors, freely correlated method factors	1151.27	40	.692	1338.72	40	.628
7	4 correlated trait factors 2 correlated method factors (semantic differential, Guttman)	243.22	40	.935	189.12	40	.947
8	4 correlated trait factors 2 correlated method factors (Likert, Guttman)	273.75	40	.927	215.86	40	.940
9	4 correlated trait factors 2 correlated method factors (Likert, semantic differential)	362.88	40	.903	309.53	40	.914

Table 4

Goodness-of-fit Indices for Comparison of Multitrait-multimethod Models

	Males (n = 412)			Females (n = 420)		
	Differences in			Differences in		
Model Comparison	χ^2	df	BBI	χ^2	df	BBI
Tests of Added Components						
Model 1 vs Model 2	3266.12	18	—	3168.35	18	—
Model 2 vs Model 3	237.39	12	.064	231.41	11	.065
Model 3 vs Model 4	54.34	2	.014	72.66	3	.020
Test of Convergent Validity						
Model 4 vs Model 5 (traits)	1684.48	18	.450	2124.78	18	.591
Tests of Discriminant Validity						
Model 4 vs Model 6 (traits)	968.57	6	.191	1212.35	6	.337
Model 4 vs Model 2 (methods)	291.73	14	.078	304.07	14	.085
Tests of Method Bias						
Model 4 vs Model 7 (Likert)	60.52	6	.016	62.75	6	.018
Model 4 vs Model 8 (semantic differential)	91.05	6	.024	89.49	6	.025
Model 4 vs Model 9 (Guttman)	180.18	6	.048	183.16	6	.051

All χ^2 difference values are statistically significant ($p < .001$).

Table 5

Factor and Error/Uniqueness Loadings, and Factor Correlations for Baseline Model Males[a]

Measure	Trait 1	Trait 2	Trait 3	Trait 4	Method I	Method II	Method III	Error/ Uniqueness
Likert								
general SC	.81* (.05)	.0	.0	.0	.58* (.12)	.0	.0	.02 (.13)
academic SC	.0	.80* (.04)	.0	.0	.05 (.05)	.0	.0	.36* (.03)
English SC	.0	.0	.80* (.04)	.0	.13 (.05)	.0	.0	.36 (.04)
mathematics SC	.0	.0	.0	.94* (.04)	.01 (.03)	.0	.0	.11* (.02)
Semantic Differential								
general SC	.66* (.05)	.0	.0	.0	.0	.57* (.07)	.0	.22* (.07)
academic SC	.0	.78* (.04)	.0	.0	.0	.41* (.06)	.0	.22* (.05)
English SC	.0	.0	.89* (.04)	.0	.0	.14 (.04)	.0	.17* (.03)
mathematics SC	.0	.0	.0	.94* (.04)	.0	.05 (.03)	.0	.12 (.02)
Guttman								
general SC	.81* (.05)	.0	.0	.0	.0	.0	.33* (.06)	.27* (.05)
academic SC	.0	.82* (.04)	.0	.0	.0	.0	.53* (.04)	.00 (.00)[b]
English SC	.0	.0	.72* (.04)	.0	.0	.0	.32* (.04)	.40* (.03)
mathematics SC	.0	.0	.0	.86* (.04)	.0	.0	.19* (.03)	.20* (.02)

Factor Correlations

Trait 1	1.0						
Trait 2	.72* (.04)	1.0					
Trait 3	.29* (.05)	.58* (.04)	1.0				
Trait 4	.42* (.05)	.74* (.03)	.24 (.05)	1.0			
Method I	.0	.0	.0	.0	1.0		
Method II	.0	.0	.0	.0	-.30* (.12)	1.0	
Method III	.0	.0	.0	.0	.67* (.14)	-.36*(.09)	1.0

[a] All values of 1.0 and .0 are fixed values.

[b] Parameter fixed to .001.

All parameter estimates differing significantly from zero are asterisked. Parenthesized values are standard errors.

SC = self-concept.

Table 6

Factor and Error/Uniqueness Loadings, and Factor Correlations for Baseline Model Females[a]

Measure	Trait 1	Trait 2	Trait 3	Trait 4	Method I	Method II	Method III	Error/ Uniqueness
Likert								
general SC	.84* (.04)	.0	.0	.0	.42* (.09)	.0	.0	.12 (.06)
academic SC	.0	.73* (.04)	.0	.0	.15 (.06)	.0	.0	.43* (.03)
English SC	.0	.0	.78* (.04)	.0	.11 (.05)	.0	.0	.39* (.04)
mathematics SC	.0	.0	.0	.93* (.04)	.01 (.04)	.0	.0	.14 (.02)
Semantic Differential								
general SC	.67* (.05)	.0	.0	.0	.0	.63* (.08)	.0	.15 (.09)
academic SC	.0	.77* (.04)	.0	.0	.0	.31* (.06)	.0	.30 (.05)
English SC	.0	.0	.85* (.04)	.0	.0	.06 (.04)	.0	.27 (.03)
mathematics SC	.0	.0	.0	.92* (.04)	.0	.06 (.03)	.0	.15 (.02)
Guttman								
general SC	.86* (.04)	.0	.0	.0	.0	.0	.31* (.06)	.18 (.04)
academic SC	.0	.75* (.05)	.0	.0	.0	.0	.62* (.05)	.00 (.00)[b]
English SC	.0	.0	.77* (.04)	.0	.0	.0	.29* (.04)	.31* (.03)
mathematics SC	.0	.0	.0	.87* (.04)	.0	.0	.19* (.03)	.18 (.02)

Factor Correlations

Trait 1	1.0						
Trait 2	.69* (.04)	1.0					
Trait 3	.36* (.05)	.65* (.04)	1.0				
Trait 4	.22* (.05)	.52* (.04)	-.01 (.05)	1.0			
Method I	.0	.0	.0	.0	1.0		
Method II	.0	.0	.0	.0	.33* (.15)	1.0	
Method III	.0	.0	.0	.0	-.87* (.11)	-.56*(.11)	1.0

[a] All values of 1.0 and .0 are fixed values.

[b] Parameter fixed to .001.

All parameter estimates differing significantly from zero are asterisked. Parenthesized values are standard errors.

SC = self-concept.

Table 7

Composite Direct Product Model Component Correlation Matrices for Males and Females (P_T)

Traits	Males				Females			
				Traits X Traits				
GSC	1.00				1.00			
ASC	1.00	1.00			1.00	1.00		
ESC	.75	.98	1.00		.67	.61	1.00	
MSC	.34	.75	.97	1.00	-.01	.17	1.00	1.00
				Methods X Methods				
LIK	1.00				1.00			
SD	.86	1.00			1.00	1.00		
GUTT	1.00	.73	1.00		1.00	.44	1.00	

GSC = general self-concept
ASC = academic self-concept
ESC = English self-concept
MSC = mathematics self-concept
LIK = Likert scale
SD = semantic differential scale
GUTT = Guttman scale