DOCUMENT RESUME

ED 333 017 TM 016 486

AUTHOR Crowley, Susan L.; Thompson, Bruce

TITLE Evaluating the Replicability of Multivariate Assessment and Evaluation Results: A Review of Various Applications of the Cross-Validation Logic.

PUB DATE 2 Apr 91

NOTE 48p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 4-6, 1991).

PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.

DESCRIPTORS *Correlation; Educational Assessment; *Evaluation Methods; Factor Analysis; Literature Reviews; *Mathematical Logic; *Multivariate Analysis; Regression (Stati tics); Reliability; Research Methodology; Statistical Significance

IDENTIFIERS Cross Validation; Invariance Principle; *Research Replication; Sampling Error

ABSTRACT

Selected methods for evaluating the stability of research results empirically are described, especially with regard to multivariate analysis. It is critically important to evaluate the influences of sampling error on obtained results; statistical significance testing does not inform judgment regarding the probable replicability or the sampling specificity of results. Data provided by K. J. Holzinger and F. Swineford (1939) are used to make the discussion more concrete. As a starting point, an example of a univariate invariance analysis is presented. Other techniques described and illustrated include: (1) multiple regression analysis; (2) factor analysis; and (3) canonical correlation analysis. Ten tables illustrate the data. An appendix contains the Statistical Package for the Social Sciences program used to implement the analyses. A 60-item list of references is included. (SLD)

holzncme.wpl 4/2/91

EVALUATING THE REPLICABILITY OF MULTIVARIATE ASSESSMENT AND EVALUATION RESULTS: A REVIEW OF VARIOUS APPLICATIONS OF THE CROSS-VALIDATION LOGIC

Susan L. Crowley Bruce Thompson

Texas A&M University 77843-4225

Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, April 6, 1991.

Abstract

Research is conducted to cumulate knowledge across studies, so assessing the generalizability of results is an essential component of scientific inquiry. Unfortunately, statistical significance testing does not inform judgment regarding the probability that results will replicate. The present paper presents a series of related logics that can be employed to empirically evaluate the replicability of results. Data reported by Holzinger and Swineford (1939), widely available to researchers and frequently used in previous heuristic examples, are employed to make the discussion concrete.

Hinkle, Wiersma and Jurs (1979, p. 415) noted that "it is becoming increasingly important for behavioral scientists to understand multivariate procedures even if they do not use them in their own research." Recent empirical studies of research practice confirm that multivariate methods are employed with some regularity in behavioral research (Elmore & Woehlke, 1988). Fish (1988) notes that there are two reasons why multivariate methods are so important in behavioral research.

First, multivariate methods limit the inflation of Type I "experimentwise" error rates. The seriousness of "experimentwise" error inflation, and what to do about it, are both matters prompting some disagreement (e.g., Bray & Maxwell, 1982, p. 343, 1985, p. 10; Hummel & Johnston, 1986). But it is clear that, "Whenever multiple statistical tests are carried out in inferential data analysis, there is a potential problem of 'probability pyramiding'" (Huberty & Morris, 1989, p. 306). And as Morrow and Frankiewicz (1979) emphasize, it is also clear that in some cases inflation of experimentwise error rates can be quite serious.

Most researchers are familiar with "testwise" alpha. But while "testwise" alpha refers to the probability of making a Type I error for a given hypothesis test, "experimentwise" error rate refers to the probability of having made a Type I error anywhere within the study, i.e., across all hypotheses. Therefore, when only one hypothesis is tested for a given group of people in a study, "experimentwise" error rate will exactly equal the "testwise" error rate.

But when more than one hypothesis is tested in a given study, the two error rates will not be equal. Witte (1985, p. 236) explains the two error rates using an intuitively appealing example involving a coin toss. If the toss of heads is equated with a Type I error, and if a coin is tossed only once, then the probability of a head on the one toss and of at least one head within the set of one toss will both equal 50%. But if the coin is tossed three times, even though the "testwise" probability of a head on each toss is 50%, the "experimentwise" probability that there will be at least one head in the whole set of three flips will be inflated to 87.5%. This dynamic is illustrated in Table 1. Analogies for research are presented in Table 2. Researchers control "testwise" error rate by picking small values, usually 0.05, for the "testwise" alpha. "Experimentwise" error rate can be limited by employing multivariate statistics.

INSERT TABLES 1 AND 2 ABOUT HERE.

Paradoxically, although the use of several univariate tests in a single study can lead to too many null hypotheses being spuriously rejected, as reflected in inflation of "experimentwise" error rate, it is also possible that the failure to employ multivariate methods can lead to a failure to identify statistically significant results which actually exist. Fish (1988) and Maxwell (1991) both provide data sets illustrating this equally disturbing possibility. Thus, "correcting" the testwise alpha level (e.g., with a Bonferroni correction--Huberty, 1987) so as to control experimentwise error rate inflation is not a satisfactory solution to this problem. The basis for this paradox is beyond the scope of the present treatment, but involves the second major reason why multivariate statistics are so important.

Multivariate methods are often vital in behavioral research because multivariate methods best honor the reality to which the researcher is purportedly trying to generalize. This is particularly important, since significance testing and error rates may not always be the most important aspect of research practice (Thompson, 1989b). Thompson (1986, p. 9) notes that the reality about which most researchers wish to generalize is usually one "in which the researcher cares about multiple outcomes, in which most outcomes have multiple causes, and in which most causes have multiple effects." Tatsuoka's (1973, p. 273) previous remarks remain telling:

> The often-heard argument, "I'm more interested in seeing how each variable, in its own right, affects the outcome" overlooks the fact that any variable taken in isolation may affect the criterion differently from the way it will act in the company of other variables. It also overlooks the fact that multivariate analysis--precisely by considering all the variables simultaneously--can throw light on how each one contributes to the relation.

Although multivariate methods have enjoyed fairly widespread usage (Thompson, 1989a; Wood & Erskine, 1976) since computers and statistical software became widely available, *multivariate methods also have been used in intriguing ways in measurement and assessment contexts*. For example, Merenda, Novack and Bonaventure (1976) reported a multivariate reliability analysis involving subtest scores from the California Test of Mental Maturity. Similarly, Sexton, McLean, Boyd, Thompson and McCormick (1988) reported results involving a multivariate concurrent validity analysis.

Unfortunately, as Nunnally (1978, p. 298) notes, "one tends to take advantage of chance in any situation where something is optimized from the data at hand." In fact, this capitalization occurs in all classical parametric methods, because all these methods (e.g., *t*-tests, ANOVA, regression, MANOVA) are least squares procedures that implicitly or explicitly (a) use weights, (b) focus on latent synthetic variables, and (c) yield effect sizes analogous to r^2, i.e., all classical analytic methods are correlational (Knapp, 1978; Thompson, 1988a).

The problem of capitalizing on sampling error when multivariate methods are used is particularly acute, because the models being tested involve a larger system of parameter estimates. For example, the problem is particularly difficult when factor analytic methods are employed, because "one has numerous possibilities for capitalizing on chance. Most extraction procedures, including principal factor solutions, reach their criterion by such capitalization. The same is true of rotational procedures, including those which rotate for simple structure"

(Gorsuch, 1983, p. 330).

Thus, it is critically important to evaluate the influences of sampling error on obtained results, i.e., the replicability or the invariance of results. Contrary to somewhat common misconceptions, statistical significance testing does not inform judgment regarding the probable replicability or the sampling-specificity of results (Carver, 1978; Thompson, 1987, 1989b). The purpose of the present paper is to describe selected methods for empirically evaluating the stability of results, especially as regards multivariate analyses.

The data reported by Holzinger and Swineford (1939, pp. 81-91), used with some frequency to illustrate multivariate statistical analyses (e.g., Gorsuch, 1983, passim; Jöreskog & Sörbom, 1986, pp. III.106-III.122), are used here to make the discussion more concrete. These data were selected for use in the examples because they are widely available, and interested readers can therefore readily replicate the analyses described here. Appendix A presents the SPSS-X program used to generate the results. Table 3 presents descriptive statistics and labels for the variables.

INSERT TABLE 3 ABOUT HERE.

A Univariate Analysis as a Starting Point

Readers more familiar with univariate analyses may appreciate an initial example of a univariate invariance analysis, prior to discussion of some methods that can be employed in the multivariate case. In a seminal article, Cohen (1968, p. 426) noted that ANOVA and ANCOVA are special cases of multiple regression analysis, and argued that in this realization "lie possibilities for more relevant and therefore more powerful exploitation of research data." Thus, regression analysis provides a good context for an invariance analysis example, because regression is so useful.

Researchers have increasingly recognized that conventional multiple regression analysis of data as they were initially collected (no conversion of intervally scaled independent variables into dichotomies or trichotomies) does not discard information or distort reality, and that the general linear model

> ...can be used equally well in experimental or non-experimental research. It can handle continuous and categorical variables. It can handle two, three, four, or more independent variables... Finally, as we will abundantly show, multiple regression analysis can do anything the analysis of variance does--sums of squares, mean squares, F ratios--and more. (Kerlinger & Pedhazur, 1973, p. 3)

Discarding variance is not generally good research practice (Kerlinger, 1986, p. 558; Thompson, 1988b) and amounts to "squandering of information" (Cohen, 1968, p. 441). As Pedhazur (1982, pp. 452-453) notes,

> Categorization of attribute variables is all too frequently resorted to in the social sciences... It is possible that some of the conflicting evidence in

the research literature of a given area may be attributed to the practice of categorization of continuous variables... Categorization leads to a loss of information, and consequently to a less sensitive analysis.

One reason why researchers may be prone to categorizing continuous variables is that some researchers unconsciously and erroneously associate ANOVA with the power of experimental designs. Humphreys (1978, p. 873) notes that:

> The basic fact is that a measure of individual differences is not an independent variable, and it does not become one by categorizing the scores and treating the categories as if they defined a variable under experimental control in a factorially designed analysis of variance.

Similarly, Humphreys and Fleishman (1974, p. 468) note that categorizing variables in a nonexperimental design using an ANOVA analysis "not infrequently produces in both the investigator and his audience the illusion that he has experimental control over the independent variable. Nothing could be more wrong."

As Cliff (1987, p. 130) notes, the practice of discarding variance on intervally scaled predictor variables to perform OVA analyses creates problems in almost all cases:

> Such divisions are not infallible; think of the persons near the borders. Some who should be highs are actually classified as lows, and vice versa. In

addition, the "barely highs" are classified the same as the "very highs," even though they are different. Therefore, reducing a reliable variable to a dichotomy makes the variable more unreliable, not less.

These various realizations have led to less frequent use of OVA methods, and to more frequent use of general linear model approaches such as regression (Elmore & Woehlke, 1988; Goodwin & Goodwin, 1985; Willson, 1982).

Example #1: Multiple Regression Analysis

The regression example utilized scores of the 301 subjects on variable T6, paragraph comprehension, as the dependent variable. Scores on variables T12, T13, T23, and T24, described in Table 3, were arbitrarily selected as the predictors for the example. The effect size in the example was noteworthy (R^2 = .277) and statistically significant (F = 28.395, df = 4/296, p < .0001).

Researchers have increasingly recognized, however, that the effect sizes in parametric analyses are subject to "shrinkage", i.e., least squares methods for data in hand tend to be overestimates of the effects that would be realized by the application of the calculated sample weights to data in other samples (Fisk, 1991). One approach to this problem invokes statistical correction formulas grounded in certain assumptions about the sample and sampling error. For example, one might employ Wherry's (1931) correction formula to R^2. The Wherry correction can be expressed as:

$$R^2 - ((1 - R^2) * (v / (n - v - 1))).$$

When applied to the example results, the corrected population estimate is:

$$
\begin{array}{l}
.27731 - ((1 - .27731) * (4 / (301 - 4 - 1))) \\
.27731 - (\quad .72269 \quad * (4 / \quad 296 \quad)) \\
.27731 - (\quad .72269 \quad * \quad .013513 \quad) \\
.27731 - \quad .009766 \\
.267543.
\end{array}
$$

However, Stevens (1986, pp. 78-84) incisively implies that researchers usually ground their work in empirical findings from previous samples, and in actual practice usually want their work to generalize to future samples in future research rather than to the unknowable population. Herzberg (1969) provides a correction for this estimate:

$$1 - ((n-1)/(n-v-1))((n-2)/(n-v-2))((n+1)/n)(1-R^2).$$

For these data the correction for R^2 would be:

$$
\begin{array}{l}
1 - (300 / 296 \quad) * (299 / 295 \quad) * (302/301) * (1-.27731) \\
1 - \quad 1.013513 \quad * \quad 1.013559 \quad * 1.003322 * \quad 0.72269 \\
1 - \quad 1.027256 \quad * 1.003322 * \quad 0.72269 \\
1 - \quad 1.030668 \quad * \quad 0.72269 \\
1 - \quad 0.744854 \\
0.255145,
\end{array}
$$

a result which even further overcorrects the estimate, and is thus still more conservative.

These sorts of statistical corrections for effect size estimates make smaller corrections both as (a) sample effect sizes are larger and (b) sample sizes are larger. However, *evaluating the stability of an effect size is a different proposition than evaluating the stability of the least squares weights used to yield the identified effect.* Since researchers usually consult the system of weights as part of their result interpretation (Thompson &

Borrello, 1985), the statistical correction formulas in and of themselves are not sufficient for evaluating the invariance of results.

Efforts to estimate the sampling specificity of coefficients for specific variables are more difficult, or at least more tedious. Some researchers randomly split their sample data, conduct separate analyses for the two subgroups, and then subjectively compare the results to determine if they appear to be similar. Two points need to be emphasized about such an approach.

First, such procedures almost always overestimate the invariance or generalizability of results, as Thompson (1984, p. 46) emphasizes. Most researchers work with samples of convenience that are homogeneous in several if not many respects, e.g., geographic location. The members of the random subgroups, then, have more in common with each other than will independent future samples drawn by other researchers. This is not said to discourage the practice of replicability analysis, but is emphasized only to give a context for the interpretation of results. It is always better to have an empirical overestimate of result replicability than to have merely a dogmatic attachment to the presumption that sample results will generalize.

Second, it is emphasized that inferences regarding replicability must be made **empirically** rather than subjectively, e.g., not by visually comparing coefficients across two randomly identified sample subgroups. Subjective comparisons will not do, because two sets of weights that appear to be different may in fact yield quite similar estimates of the synthetic or latent variables actually being correlated in all parametric analyses, e.g., the synthetic variable $\hat{Y}$ being correlated with Y to yield R in a regression analysis. Rowell (1991) provides a concrete example of just such a case. Cliff (1987, pp. 177-178) suggests that such cases involve "insensitivity" of the weights to departures from least squares constraints.

Cross-validation is one vehicle for empirically exploring the stability of regression weights and resulting effect sizes across samples (Huck, Cormier & Bounds, 1974, pp. 159-160). Rowell (1991) and Thompson (1989b) provide examples for the regression case. Of course, other empirical methods are available for evaluating result replicability, such as the *bootstrap* (Diaconis & Efron, 1983; Lunneborg, 1987). Thompson and Melancon (1990) provide an example of the application of the bootstrap in the regression case.

In the first step of the cross-validation process the sample is divided into two subsamples, usually randomly. However, for the purposes of the present heuristic example, the two schools represented in the Holzinger and Swineford (1939) data were used as the subgroups.

Next, separate analyses were conducted for the two data sets. The effect size for the 156 students at Pasteur School, $R^2 = .240$, was statistically significant ($\underline{F} = 11.927$, $\underline{df} = 4/151$, $\underline{p} < .0001$). The prediction equation was:

$$Z_Y \leftarrow\!\!\text{----}\; \hat{Y}_{11} = (+.032125 * Z_{T12}) + (-.031195 * Z_{T13}) + (+.383641 * Z_{T23}) + (+.170181 * Z_{T24})$$

The effect size for the 145 students at Grant-White School, R^2 = .313, was statistically significant (F = 15.979, df = 4/140, p < .0001). The prediction equation was:

$$Z_Y \text{ <---- } \hat{Y}_{22} = (-.252179 * Z_{T12}) + (+.215562 * Z_{T13}) + (+.295226 * Z_{T23}) + (+.333045 * Z_{T24})$$

The third step in the analysis requires that the predictor variables in each subsample be standardized into z-score form using each group's own means and SDs, and then that new $\hat{Y}$'s be calculated for subsample one using group two's weights (called here $\hat{Y}_{12}$), and for subsample two using group one's weights (called here $\hat{Y}_{21}$). Once this is done, correlation coefficients are computed among the various synthetic variable estimates. For the example data, these results are presented in Table 4.

INSERT TABLE 4 ABOUT HERE.

The Table 4 invariance coefficients (.8506 and .8613) suggest that the subsample results are reasonably invariant. The results also illustrate the importance of empirically evaluating invariance, since the weights yield reasonably comparable estimates of $\hat{Y}$, notwithstanding the fact that the beta weights might appear different on the basis of subjective inspection (e.g., +.032125 vs -.252179, -.031195 vs +.215562).

The researcher with such results will conclude that the R^2 effect size is relatively stable, that the beta weights fluctuate, but that the effect sizes tend to be "insensitive" to these fluctuations. Of course, the regression results for the full sample will then provide the final basis for interpretation. The subsample results are employed to evaluate result stability, and are not used as the basis for interpretation. The results for the full sample are used for interpretation, since these results should theoretically be the most stable, as a function of sample size.

Example Invariance Analyses for Two Multivariate Analyses

The same logic for evaluating invariance can be readily generalized to analyses that are multivariate. Factor analysis and canonical correlation analysis were the two methods selected as examples of this generalization.

Example #2: Factor Analysis

Factor analysis has been closely associated with evaluating the construct validity of measures. Nunnally (1978, p. 111) notes that "construct validity has been spoken of as 'trait validity' and 'factorial validity.'" Gorsuch (1983, pp. 350-351) suggests that

> A prime use of factor analysis has been in the development of both the theoretical constructs for an area and the operational representatives for the theoretical constructs... If a theory has clearly defined constructs, then scales can be directly built to embody those constructs.

Thus, "factor analysis is intimately involved with questions of validity... Factor analysis is at the heart of the measurement of psychological constructs" (Nunnally, 1978, p. 112).

Twenty-four variables from the Holzinger and Swineford (1939) data, T1 through T24, were employed in this example. A variety of invariance logics can be applied in factor analysis, including "best fit" Procrustean rotation across sample splits (Thompson, 1991), bootstrap factor analysis (Thompson, 1988c), and various other methods described and compared by Guadagnoli and Velicer (1988). However, the method selected for discussion here is a generalization of the regression cross-validation strategy, a method familiar to many researchers.

A variety of procedures can be employed to calculate the synthetic variables in factor analysis, called factor scores (Thompson, 1983). However, the most common estimation procedure is the regression procedure represented by the matrix algebra algorithm:

$$\mathbf{Z}_{NxV}\ \mathbf{W}_{VxF},$$

where

$$\mathbf{W}_{VxF} = \mathbf{R}_{VxV}^{-1}\ \mathbf{P}_{VxF},$$

and where $\mathbf{Z}$ is the set of $\underline{v}$ z-scores for each of the $\underline{n}$ subjects, $\mathbf{R}^{-1}$ is the inverse of the correlation matrix for the variable set, and $\mathbf{P}$ is the orthogonal factor pattern/structure matrix.

The **W** matrix is related to the beta weights employed to estimate the synthetic variable scores in regression. Thus, one approach to evaluating invariance of factor analytic results could invoke a comparison of the synthetic factor scores derived using **W** matrices across sample splits.

Table 5 presents the varimax-rotated structure matrix for the 156 Pasteur students, and Table 6 presents the associated Weight matrix for the students. Table 7 presents the varimax-rotated structure matrix for the 145 Grant-White students, and Table 8 presents the associated Weight matrix for the students.

INSERT TABLES 5 THROUGH 8 ABOUT HERE.

Appendix A presents the SPSS-X commands employed to compute the factor scores for each group involving each group's own data and each group's own weights, i.e., variables "fs111" through "fs114" for the 156 Pasteur students, and variables "fs221" through "fs224" for the 145 Grant-White students. Appendix A also presents the SPSS-X commands employed to compute the factor scores for each group involving each group's own data and *the other group's weights*, i.e., variables "fs121" through "fs124" for the 156 Pasteur students, and variables "fs211" through "fs214" for the 145 Grant-White students.

Table 9 presents the invariance coefficients for analysis, and these coefficients are **bolded**. The first set of coefficients for the four principal components scores is: .9677, .9447, .9740, and .9590. The second set of coefficients is: .9695, .9606, .9565, .9633. If such results had occurred across random subgroups of a sample, the researcher would doubtless be relatively sanguine about the stability of results.

INSERT TABLE 9 ABOUT HERE.

Example #3: Canonical Correlation Analysis

Though multiple regression is a useful analytic method, canonical correlation analysis, and not regression analysis, is the most general case of the general linear model (Baggaley, 1981, p. 129; Fornell, 1978, p. 168). In an important article, Knapp (1978, p. 410) demonstrated this in some mathematical detail and concluded that "virtually all of the commonly encountered tests of significance can be treated as speci.l cases of canonical correlation analysis." Thompson (1988a) illustrates how canonical correlation analysis can be employed to implement all the parametric tests that canonical methods subsume as special cases.

Thus, canonical correlation analysis is a powerful analytic paradigm that can be applied to myriad research problems. The method is valuable because it honors the complexity of reality by simultaneously considering all relationships among variables, and does not require that intervally scaled predictor variables be converted to nominal scale. As Stevens (1986, p. 373, emphasis omitted) notes,

> canonical correlation... is appropriate if the wish is to parsimoniously describe the number and nature of mutually independent relationships between the two [or more] sets [of variables]... Since the combinations [of the variables derived through least squares weighting] are uncorrelated, one will obtain a very nice additive partitioning of the total between association.

Canonical analysis yields an effect size estimate, R_C^2, that is akin to the R^2 effect in multiple regression. L.ke regression, which yields both weights (called beta weights) and correlations coefficients (called **structure coefficients**) between observed predictors (e.g., X_{T12}) and the synthetic variable, $\hat{Y}$ (Thompson & Borrello, 1985), canonical analysis also yields least squares weights (called canonical function coefficients) and structure coefficients. The two sets of coefficients have the same function and meaning in canonical analysis as they do in regression, and are the two primary rivals for evaluating variable importance at the function level (Harris, 1989; Kerlinger & Pedhazur, 1973, p. 344; Levine, 1977, p. 20, Meredith, 1964, p. 55; Thompson, in press-b).

However, though canonical effect sizes tend to be reasonably stable across samples (Thompson, 1990), the individual function coefficients and structure coefficients that are an important component of the analysis tend to be less stable (Thompson, in press-a). Thus, invariance analyses are very important in the canonical case, and several methods can be utilized (Thompson, 1984). The single method illustrated here is in the same genre as the previous examples.

Variables T1 through T4 were related to variables T5 through T9 in the present example. Four uncorrelated canonical functions were possible in the example, since the smallest variable set consisted of four variables. For the full sample the effect size (R_C^2 = .00720) for the third function was negligible and the likelihood ratio for roots three and four was not statistically significant, therefore invariance analyses were conducted only for the first two functions. Appendix A presents the COMPUTE statements with the canonical function weights used to calculate the synthetic variables correlated in the analysis. For example, the synthetic criterion composite variable on Function I for the 145 Pasteur students using their data and their weights ("CRIT111") was:

$$crit111 = (.90048 * p_{ZT1}) + (.09741 * p_{ZT2}) + (.05615 * p_{ZT3}) + (.08840 * p_{ZT4})$$

Table 10 presents the "shrunken" effect size coefficients derived for the example. For the first subsample, the R_C of .5108 shrinks to .3091 when group two's weights are applied to group one's data, and the second R_C of .2788 shrinks to -.1831. For the second subsample, the R_C of .5195 shrinks to .3685 when group one's weights are applied to group two's data, and the second R_C of .1679 shrinks to -.1282. Though the effect sizes for both the two functions were relatively similar across the sample splits, the weights employed in the analysis were appreciably more comparable for the first than for the second function in both subsamples.

INSERT TABLE 10 ABOUT HERE.

Summary

Statistical significance testing does not inform the researcher regarding the replicability of results. Yet the business of science is formulating **generalizable** insight. No one study, taken singly, establishes the basis for such insight. As Neale and Liebert (1986, p. 290) observe:

> No one study, however shrewdly designed and carefully executed, can provide convincing support for a causal hypothesis or theoretical statement...

> Too many possible (if not plausible) confounds, limitations on generality, and alternative interpretations can be offered for any one observation. Moreover, each of the basic methods of research (experimental, correlational, and case study) and techniques of comparison (within- or between-subjects) has intrinsic limitations. How, then, does social science theory advance through research? The answer is, by collecting a diverse body of evidence about any major theoretical proposition.

Evaluating the generalizability of multivariate results to other samples of subjects or of variables is a daunting task, but a task which the serious scholar can ill-afford to shirk. Science will cumulate knowledge only to the extent that idiosyncratic findings are recognized as such, and significance testing is not particularly useful for making this evaluation. The present paper has illustrated the application of a few of the various logics available to the researcher who wishes to pursue such investigations.

References

Baggaley, A. R. (1981). Multivariate analysis: An introduction for consumers of behavioral research. Evaluation Review, 5, 123-131.

Bray, J.H., & Maxwell, S.E. (1982). Analyzing and interpreting significant MANOVAs. Review of Educational Research, 52, 340-367.

Bray, J.H., & Maxwell, S.E. (1985). Multivariate analysis of variance. Newbury Park: SAGE.

Carver, R. P. (1978). The case against statistical significance testing. Harvard Educational Review, 48, 378-399.

Cliff, N. (1987). Analyzing multivariate data. San Diego: Harcourt Brace Jovanovich.

Cohen, J. (1968). Multiple regression as a general data-analytic system. Psychological Bulletin, 70, 426-443.

Diaconis, P., & Efron, B. (1983). Computer-intensive methods in statistics. Scientific American, 248(5), 116-130.

Elmore, P.B., & Woehlke, P.L. (1988). Statistical methods employed in American Educational Research Journal, Educational Researcher, and Review of Educational Research from 1978 to 1987. Educational Researcher, 17(9), 19-20.

Fish, L.J. (1988). Why multivariate methods are usually vital. Measurement and Evaluation in Counseling and Development, 21, 130-137.

Fisk, Y.H. (1991, April). Various approaches to effect size estimation. Paper presented at the annual meeting of the

American Educational Research Association, Chicago.

Fornell, C. (1978). Three approaches to canonical analysis. Journal of the Market Research Society, 20, 166-181.

Goodwin, L.D., & Goodwin, W.L. (1985). Statistical techniques in AERJ articles, 1979-1983: The preparation of graduate students to read the educational research literature. Educational Researcher, 14(2), 5-11.

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

Guadagnoli, E., & Velicer, W. (1988, August). A comparison of pattern matching indices. Paper presented at the annual meeting of the American Psychological Association, Atlanta.

Harris, R.J. (1989). A canonical cautionary. Multivariate Behavioral Research, 24, 17-39.

Herzberg, P.A. (1969). The parameters of cross validation. Psychometrika, Monograph supplement, No. 16.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1979). Applied statistics for the behavioral sciences. Chicago: Rand McNally.

Holzinger, K.J., & Swineford, F. (1939). A study in factor analysis: The stability of a bi-factor solution (No. 48). Chicago, IL: University of Chicago.

Huberty, C.J. (1987). On statistical testing. Educational Researcher, 16(8), 4-9.

Huberty, C.J, & Morris, J.D. (1989). Multivariate analysis versus multiple univariate analysis. Psychological Bulletin, 105, 302-308.

Huck, S., Cormier, W., & Bounds, W. (1974). Reading statistics and research. New York: Harper & Row.

Hummel, T.J., & Johnston, C.B. (1986, April). An empirical comparison of size and power of seven methods for analyzing multivariate data in the two-sample case. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 269 480)

Humphreys, L.G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. Journal of Educational Psychology, 70, 873-876.

Humphreys, L.G., & Fleishman, A. (1974). Pseudo-orthogonal and other analysis of variance designs involving individual-differences variables. Journal of Educational Psychology, 66, 464-472.

Jöreskog, K.G., & Sörbom, D. (1986). LISREL 6: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods (4th ed.). Mooresville, IN: Scientific Software.

Kerlinger, F. N. (1986). Foundations of behavioral research (3rd ed.). New York: Holt, Rinehart and Winston.

Kerlinger, F. N., & Pedhazur, E. J. (1973). Multiple regression in behavioral research. New York: Holt, Rinehart and Winston.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. Psychological Bulletin, 85, 410-416.

Levine, M. S. (1977). Canonical analysis and factor comparison. Beverly Hills: SAGE.

Lunneborg, C.E. (1987). Bootstrap applications for the behavioral sciences. Seattle: University of Washington.

Maxwell, S. (1991). Recent developments in MANOVA applications. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2, pp. 137-168). Greenwich, CT: JAI Press.

Meredith, W. (1964). Canonical correlations with fallible data. Psychometrika, 29, 55-65.

Merenda, P., Novack, H., & Bonaventure, E. (1976). Multivariate analysis of the California test of mental maturity, primary forms. Psychological Reports, 38, 487-493.

Morrow, J.R., Jr., & Frankiewicz, R.G. (1979). Strategies for the analysis of repeated and multiple measures designs. Research Quarterly, 50, 297-304.

Neale, J.M., & Liebert, R.M. (1986). Science and behavior: An introduction to methods of research (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

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

Pedhazur, E. J. (1982). Multiple regression in behavioral research: Explanation and prediction (2nd ed.). New York: Holt, Rinehart and Winston.

Rowell, R.K. (1991, April). Double cross-validation in multiple regression: A method of estimating the stability of results. Paper presented at the annual meeting of the American

Educational Research Association, Chicago.

Sexton, J.D., McLean, M., Boyd, R.D., Thompson, B., & McCormick, K. (1988). Criterion-related validity of a new standardized developmental measure for use with infants who are handicapped. Measurement and Evaluation in Counseling and Development, 21, 16-24.

Stevens, J. (1986). Applied multivariate statistics for the social sciences. Hillsdale, NJ: Erlbaum.

Tatsuoka, M. M. (1973). Multivariate analysis in educational research. In F. N. Kerlinger (Ed.), Review of research in education (pp. 273-319). Itasca, IL: Peacock.

Thompson, B. (1983, January). The calculation of factor scores: an alternative. Paper presented at the annual meeting of the Southwest Educational Research Association, Houston. [Order document #04501 from National Auxiliary Publication Service, P.O. Box 3513, Grand Central Station, NY, NY 10163]

Thompson, B. (1984). Canonical correlation analysis: Uses and interpretation. Newbury Park: SAGE.

Thompson, B. (1986, November). Two reasons why multivariate methods are usually vital. Paper presented at the annual meeting of the Mid-South Educational Research Association, Memphis.

Thompson, B. (1987, April). The use (and misuse) of statistical significance testing: Some recommendations for improved editorial policy and practice. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C. (ERIC Document Reproduction Service No. ED 287

868)

Thompson, B. (1988a, April). Canonical correlation analysis: An explanation with comments on correct practice. Paper presented at the annual meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 295 957)

Thompson, B. (1988b). Discarding variance: A cardinal sin in research. Measurement and Evaluation in Counseling and Development, 21, 3-4.

Thompson, B. (1988c). Program FACSTRAP: A program that computes bootstrap estimates of factor structure. Educational and Psychological Measurement, 43, 681-686.

Thompson, B. (1989a, August). Applications of multivariate statistics: A bibliography of canonical correlation analysis studies. Paper presented that the annual meeting of the American Psychological Association, New Orleans. (ERIC Reproduction Service No. ED 311 070)

Thompson, B. (1989b). Statistical significance, result importance, and result generalizability: Three noteworthy but somewhat different issues. Measurement and Evaluation in Counseling and Development, 22, 2-6.

Thompson, B. (1990). Finding a correction for the sampling error in multivariate measures of relationship: A Monte Carlo study. Educational and Psychological Measurement, 50, 15-31.

Thompson, B. (1991). A partial test distribution for cosines among factors across samples. In B. Thompson (Ed.), Advances in social science methodology (Vol. 2, pp. 81-97). Greenwich, CT:

JAI Press.

Thompson, B. (in press-a). Invariance of multivariate results. Journal of Experimental Education.

Thompson, B. (in press-b). A primer on the logic and use of canonical correlation analysis. Measurement and Evaluation in Counseling and Development.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. Educational and Psychological Measurement, 45, 203-209.

Thompson, B., & Melancon, J.G. (1990, November). Bootstrap versus statistical effect size corrections: A comparison with data from the Finding Embedded Figures Test. Paper presented at the annual meeting of the Mid-South Educational Research Association, New Orleans.

Wherry, R.J. (1931). A new formula for predicting the shrinkage of the coefficient of multiple correlation. Annals of Mathematical Statistics, 2, 440-451.

Willson, V. L. (1982, February). Misuses of regression approaches to ANOVA and ANCOVA. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin. (ERIC Document Reproduction Service No. ED 222 522)

Witte, R.S. (1985). Statistics (2nd ed.). New York: Holt, Rinehart and Winston.

Wood, D.A., & Erskine, J.A. (1976). Strategies in canonical correlation with application to behavioral data. Educational and Psychological Measurement, 36, 861-878.

Table 1
All Possible Families of Outcomes
for a Fair Coin Flipped Three Times

	Flip # 1	2	3
1.	T	T	T
2.	H	T	T
3.	T	H	T
4.	T	T	H
5.	H	H	T
6.	H	T	H
7.	T	H	H
8.	H	H	H
p of H on each Flip	50%	50%	50%

(for rows 2–8) p of 1 or more H's (TW error analog) in set of 3 Flips = 7/8 = 87.5%

or

where TW error analog = .50,
EW $p = 1 - (1 - .5)^3$
$= 1 - .5^3 = 1 - .125 = .875$

Note. The probability of one or more occurrences of a given outcome in a set of events is $1 - (1-p)^k$, where p is the probability of the given occurrence on each trial and k is the number of trials in a set of perfectly independent events.

Table 2
Formula for Estimating Experimentwise Type I Error Inflation
When Hypotheses are Perfectly Uncorrelated

	TW alpha		Tests		Experimentwise alpha
1 - (1 -	0.05)	**	1	=	
1 - (	0.95)	**	1	=	[a]
1 -	0.95			=	0.05000
Range Over Testwise (TW) alpha = .01					
1 - (1 -	0.01)	**	5	=	0.04901
1 - (1 -	0.01)	**	10	=	0.09562
1 - (1 -	0.01)	**	20	=	0.18209
Range Over Testwise (TW) alpha = .05					
1 - (1 -	0.05)	**	5	=	0.22622
1 - (1 -	0.05)	**	10	=	0.40126
1 - (1 -	0.05)	**	20	=	0.64151
Range Over Testwise (TW) alpha = .10					
1 - (1 -	0.10)	**	5	=	0.40951
1 - (1 -	0.10)	**	10	=	0.65132
1 - (1 -	0.10)	**	20	=	0.87842

Note. "**" = "raise to the power of".

[a]These calculations are presented (a) to illustrate the implementation of the formula step by step and (b) to demonstrate that when only one test is conducted, the experimentwise error rate equals the testwise error rate, as should be expected if the formula behaves properly.

Table 3
Descriptive Statistics for Holzinger and Swineford (1939) Data

Variable	Mean	SD	Variable Label
SCHOOL	1.482	.500	
T1	29.615	7.005	VISUAL PERCEPTION TEST FROM SPEARMAN VPT
T2	24.352	4.710	CUBES, SIMPLIFICATION OF BRIGHAM'S SPATI
T3	14.229	2.830	PAPER FORM BOARD--SHAPES THAT CAN BE COM
T4	18.003	9.048	LOZENGES FROM THORNDIKE--SHAPES FLIPPED
T5	40.591	12.381	GENERAL INFORMATION VERBAL TEST
T6	9.183	3.492	PARAGRAPH COMPREHENSION TEST
T7	17.362	5.162	SENTENCE COMPLETION TEST
T8	26.126	5.675	WORD CLASSIFICATION--WHICH WORD NOT BELO
T9	15.299	7.669	WORD MEANING TEST
T10	96.276	25.059	SPEEDED ADDITION TEST
T11	69.163	15.670	SPEEDED CODE TEST--TRANSFORM SHAPES INTO
T12	110.542	20.252	SPEEDED COUNTING OF DOTS IN SHAPE
T13	193.468	36.329	SPEEDED DISCRIM STRAIGHT AND CURVED CAPS
T14	175.153	11.508	MEMORY OF TARGET WORDS
T15	90.010	7.729	MEMORY OF TARGET NUMBERS
T16	102.525	7.633	MEMORY OF TARGET SHAPES
T17	8.233	4.916	MEMORY OF OBJECT-NUMBER ASSOCIATION TARG
T18	9.425	4.488	MEMORY OF NUMBER-OBJECT ASSOCIATION TARG
T19	14.037	4.077	MEMORY OF FIGURE-WORD ASSOCIATION TARGET
T20	26.890	19.334	DEDUCTIVE MATH ABILITY
T21	14.249	4.562	MATH NUMBER PUZZLES
T22	26.239	9.197	MATH WORD PROBLEM REASONING
T23	18.136	9.140	COMPLETION OF A MATH NUMBER SERIES
T24	24.266	4.735	WOODY-MCCALL MIXED MATH FUNDAMENTALS TES
T25	15.648	3.086	REVISION OF T3--PAPER FORM BOARD
T26	36.303	8.339	FLAGS--POSSIBLE SUBSTITUTE FOR T4 LOZENG

Table 4
Invariance Coefficients for the Regression Example

	Y	$\hat{Y}_{11}$	$\hat{Y}_{12}$	$\hat{Y}_{21}$	$\hat{Y}_{22}$
Y	1.0000				
$\hat{Y}_{11}$	.4900[a]	1.0000			
$\hat{Y}_{12}$	.4168[c]	.8506[e]	1.0000		
$\hat{Y}_{21}$	.4822[d]	.	.	1.0000	
$\hat{Y}_{22}$	.5599[b]	.	.	.8613[f]	1.0000

[a]This is the multiple R for the 156 students at Pasteur.
[b]This is the multiple R for the 145 students at Grant-White.
[c]This is the "shrunken" value of the R for the 156 students at Pasteur, based on using the beta weights for the Grant-White students. The shrinkage is $.4900^2 - .4168^2 = .2401 - .1737 = .0664$.
[d]This is the "shrunken" value of the R for the 145 students at Grant-White, based on using the beta weights for the Pasteur students. The shrinkage is $.5599^2 - .4822^2 = .3135 - .2325 = .0810$.
[e]This is the invariance coefficient for the 156 students at Pasteur.
[f]This is the invariance coefficient for the 145 students at Grant-White.

Table 5
Varimax Rotated Pattern/Structure Matrix for Pasteur Students
(n = 156)

Variable	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
T1	.35496	.62024	.08623	.13566
T2	.01632	.64395	-.06319	-.06295
T3	.09557	.55298	.05815	-.24556
T4	-.01989	.72448	.14223	.19744
T5	.84834	.01298	.14422	-.07502
T6	.83010	.13268	.13814	.09395
T7	.89409	.06204	.06757	-.01908
T8	.75202	.15203	.08439	.14611
T9	.80579	.22468	.17590	.04170
T10	.16125	-.21005	.75728	.11505
T11	.35796	.03703	.62533	.20684
T12	.03551	.08427	.71847	-.05683
T13	.01455	.27829	.61539	.03179
T14	.03530	.05151	.00034	.74889
T15	-.14516	.15742	-.08054	.71260
T16	.08347	.41132	.26089	.50423
T17	.05943	-.09097	.33372	.59274
T18	.14625	-.06880	.03472	.61530
T19	.06763	.22600	.27591	.43278
T20	.11427	.60330	.04047	.22728
T21	.29862	.41483	.43691	.13297
T22	.50750	.48108	.16256	.01041
T23	.39696	.58608	.24882	.15038
T24	.25074	.27372	.53294	.23170

Table 6
Regression Factor Score Weight Matrix for Pasteur Students
(n = 156)

Variable	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
T1	.03613	.18800	-.05855	.00734
T2	-.05198	.25519	-.06732	-.06730
T3	-.03485	.21754	.00877	-.15560
T4	-.09427	.25735	-.00020	.01818
T5	.23775	-.08361	-.02595	-.04572
T6	.22213	-.04758	-.05440	.02127
T7	.25641	-.06830	-.07813	-.01511
T8	.20342	-.03259	-.07813	.04886
T9	.20066	-.00839	-.03323	-.01324
T10	-.02154	-.15523	.34434	-.02159
T11	.03008	-.07765	.23000	.01804
T12	-.08539	-.01425	.34184	-.11566
T13	-.09797	.06293	.27168	-.07840
T14	.01041	-.03819	-.09559	.32932
T15	-.04524	.02857	-.11519	.31381
T16	-.04415	.09627	.02875	.16406
T17	-.00964	-.10607	.08931	.23123
T18	.05097	-.08815	-.06895	.27537
T19	-.03277	.03066	.05849	.14391
T20	-.02896	.20182	-.06092	.05311
T21	-.00314	.09159	.13054	-.02052
T22	.08612	.12328	-.01656	-.04619
T23	.03258	.17888	.01494	-.00248
T24	-.01461	.02850	.18016	.02203

Table 7
Varimax Rotated Pattern/Structure Matrix for Grant-White Students
(n = 145)

Variable	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
T1	.18969	.69323	.17062	.16118
T2	.09441	.65983	.02583	.01294
T3	.16318	.51715	.13603	.12246
T4	.24833	.71458	.06545	.07524
T5	.77924	.21503	.19557	.07616
T6	.80685	.15161	.07417	.19990
T7	.84695	.11903	.16615	.07617
T8	.65732	.25086	.25931	.11186
T9	.84512	.16299	.04717	.18076
T10	.16110	-.07206	.83716	.11438
T11	.17916	.12677	.60857	.33076
T12	.02526	.22914	.78476	.04228
T13	.19371	.45332	.61470	.02085
T14	.20665	.02488	.05357	.66968
T15	.07807	.16183	.01560	.64922
T16	.07075	.45477	.03646	.57780
T17	.16902	-.07154	.27888	.70520
T18	.00698	.35969	.36721	.47717
T19	.17820	.13186	.22208	.45047
T20	.45866	.38635	.05303	.33040
T21	.17998	.48808	.44788	.18824
T22	.44465	.36018	.10038	.33810
T23	.44543	.48780	.20782	.23026
T24	.41133	.08011	.55939	.28566

Table 8
Regression Factor Score Weight Matrix for Grant-White Students
(n = 145)

Variable	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
T1	-.05942	.27437	-.02782	-.03557
T2	-.06284	.30421	-.06587	-.08359
T3	-.03719	.20040	-.01860	-.02880
T4	-.02332	.29744	-.07279	-.07783
T5	.23536	-.03608	-.00039	-.09318
T6	.25056	-.07273	-.06840	-.00593
T7	.27651	-.08955	-.00928	-.08556
T8	.17620	-.01198	.03046	-.07281
T9	.26728	-.06760	-.08271	-.01840
T10	-.00634	-.15925	.37219	-.05343
T11	-.03412	-.06617	.21552	.06830
T12	-.08810	.02109	.33573	-.10645
T13	-.04231	.12665	.22256	-.14094
T14	-.00362	-.10132	-.07867	.33244
T15	-.06587	-.00971	-.09831	.32409
T16	-.10440	.14043	-.10957	.24912
T17	-.02489	-.17071	.03516	.33842
T18	-.12885	.08030	.07278	.17023
T19	-.01861	-.03707	.01905	.18221
T20	.07507	.07865	-.09283	.07819
T21	-.05779	.14348	.12063	-.03113
T22	.06915	.06203	-.06786	.08093
T23	.05464	.12554	-.01729	-.00647
T24	.06746	-.11178	.18538	.03025

Table 9
Factor Score Correlation Matrix

	FS111	FS112	FS113	FS114	FS211	FS212	FS213	FS214	FS221	FS222	FS223
FS111	1.0000 (156)										
FS112	.0000 (156)	1.0000 (156)									
FS113	.0000 (156)	.0000 (156)	1.0000 (156)								
FS114	.0000 (156)	.0000 (156)	.0000 (156)	1.0000 (156)							
FS211	. (0)	. (0)	. (0)	. (0)	1.0000 (145)						
FS212	. (0)	. (0)	. (0)	. (0)	.2397 (145)	1.0000 (145)					
FS213	. (0)	. (0)	. (0)	. (0)	.0474 (145)	.0790 (145)	1.0000 (145)				
FS214	. (0)	. (0)	. (0)	. (0)	.2756 (145)	.2015 (145)	.1247 (145)	1.0000 (145)			
FS221	. (0)	. (0)	. (0)	. (0)	**.9677** (145)	.1956 (145)	-.0006 (145)	.1380 (145)	1.0000 (145)		
FS222	. (0)	. (0)	. (0)	. (0)	.0747 (145)	**.9447** (145)	.1017 (145)	.0734 (145)	.0000 (145)	1.0000 (145)	
FS223	. (0)	. (0)	. (0)	. (0)	.0518 (145)	-.0497 (145)	**.9740** (145)	.0694 (145)	.0000 (145)	.0000 (145)	1.0000 (145)
FS224	. (0)	. (0)	. (0)	. (0)	.1281 (145)	.1401 (145)	.0808 (145)	**.9590** (145)	.0000 (145)	.0000 (145)	.0000 (145)
FS121	**.9695** (156)	.0626 (156)	.0248 (156)	-.1319 (156)	. (0)	. (0)	. (0)	. (0)	. (0)	. (0)	. (0)
FS122	-.0520 (156)	**.9606** (156)	-.0241 (156)	-.0578 (156)	. (0)	. (0)	. (0)	. (0)	. (0)	. (0)	. (0)
FS123	.0844 (156)	-.2062 (156)	**.9565** (156)	-.0161 (156)	. (0)	. (0)	. (0)	. (0)	. (0)	. (0)	. (0)
FS124	-.1401 (156)	.0399 (156)	.0234 (156)	**.9633** (156)	. (0)	. (0)	. (0)	. (0)	. (0)	. (0)	. (0)

Table 9 (cont.)

	FS224	FS121	FS122	FS123	FS124
FS121	.	1.0000	-.0236	.0781	-.2568
	(0)	(156)	(156)	(156)	(156)
FS122	.	-.0236	1.0000	-.2031	-.0436
	(0)	(156)	(156)	(156)	(156)
FS123	.	.0781	-.2031	1.0000	-.0367
	(0)	(156)	(156)	(156)	(156)
FS124	.	-.2568	-.0436	-.0367	1.0000
	(0)	(156)	(156)	(156)	(156)

Note. The two sets of four invariance coefficients are presented in **bold**.

Table 10
Correlation Coefficients for Various Pairs of Canonical Variate Scores

	CRIT111	CRIT112	PRED111	PRED112	CRIT121	CRIT122	PRED121	PRED122	CRIT211	CRIT212	PRED211
CRIT111	1.0000										
	(156)										
CRIT112	.0000	1.0000									
	(156)	(156)									
PRED111	*[.5108]*	.0000	1.0000								
	(156)	(156)	(156)								
PRED112	.0000	*[.2788]*	.0000	1.0000							
	(156)	(156)	(156)	(156)							
CRIT121	.8367	-.2956	.4274	-.0824	1.0000						
	(156)	(156)	(156)	(156)	(156)						
CRIT122	.0333	-.5702	.0170	-.1590	-.0082	1.0000					
	(156)	(156)	(156)	(156)	(156)	(156)					
PRED121	.4264	.0993	.8348	.3563	.3091	-.0211	1.0000				
	(156)	(156)	(156)	(156)	(156)	(156)	(156)				
PRED122	-.1180	.2471	-.2311	.8864	-.1465	-.1831	.0035	1.0000			
	(156)	(156)	(156)	(156)	(156)	(156)	(156)	(156)			
CRIT211	.	.	.	.	.	.	.	.	1.0000		
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(145)		
CRIT212	.	.	.	.	.	.	.	.	-.0310	1.0000	
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(145)	(145)	
PRED211	.	.	.	.	.	.	.	.	.3685	-.1174	1.0000
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(145)	(145)	(145)
PRED212	.	.	.	.	.	.	.	.	.1632	-.1282	.0060
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(145)	(145)	(145)
CRIT221	.	.	.	.	.	.	.	.	.8306	-.3105	.4312
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(145)	(145)	(145)
CRIT222	.	.	.	.	.	.	.	.	.0769	-.6145	-.0328
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(145)	(145)	(145)
PRED221	.	.	.	.	.	.	.	.	.4315	-.1613	.8301
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(145)	(145)	(145)
PRED222	.	.	.	.	.	.	.	.	.0129	-.1032	-.1951
	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(0)	(145)	(145)	(145)

Table 10 (cont.)

	PRED212	CRIT221	CRIT222	PRED221	PRED222
PRED212	1.0000	.1728	.1492	**.3326**	**.8887**
	(145)	(145)	(145)	(145)	(145)
CRIT221	.1728	1.0000	.0000	*[.5195]*	.0000
	(145)	(145)	(145)	(145)	(145)
CRIT222	.1492	.0000	1.0000	.0000	*[.1679]*
	(145)	(145)	(145)	(145)	(145)
PRED221	.3326	.5195	.0000	1.0000	.0000
	(145)	(145)	(145)	(145)	(145)
PRED222	.8887	.0000	.1679	.0000	1.0000
	(145)	(145)	(145)	(145)	(145)

Note. The two R_c values for the two subsamples are presented in *[square brackets in italics]*. The two sets of invariance coefficients are presented in **bold**.

Appendix A:
SPSS-X Program Used to Implement the Reported Analyses

```
TITLE 'Holzinger & Swineford (1939) Data **Citation in Comment**'
C          O          M          M          E          N          T
**********************************************************************
COMMENT  Holzinger, K.J., & Swineford, F. (1939). A study in factor analysis:
COMMENT     The stability of a bi-factor solution (No. 48). Chicago, IL:
COMMENT     University of Chicago. (data on pp. 81-91)
C          O          M          M          E          N          T
**********************************************************************
DATA LIST FILE=BT RECORDS=2
  /1 ID 1-3 SEX 4 AGEYR 6-7 AGEMO 8-9
  T1 11-12 T2 14-15 T3 17-18 T4 20-21 T5 23-24 T6 26-27
  T7 29-30 T8 32-33 T9 35-36 T10 38-40 T11 42-44 T12 46-48
  T13 50-52 T14 54-56 T15 58-60 T16 62-64 T17 66-67
  T18 69-70 T19 72-73 T20 74-76 T21 78-79
  /2 T22 11-12 T23 14-15 T24 17-18
  T25 20-21 T26 23-24
COMPUTE SCHOOL=1
IF (ID GT 200)SCHOOL=2
IF (ID GE 1 AND ID LE 85)GRADE=7
IF (ID GE 86 AND ID LE 168)GRADE=8
IF (ID GE 201 AND ID LE 281)GRADE=7
IF (ID GE 282 AND ID LE 351)GRADE=8
IF (ID GE 1 AND ID LE 44)TRACK=2
IF (ID GE 45 AND ID LE 85)TRACK=1
IF (ID GE 86 AND ID LE 129)TRACK=2
IF (ID GE 130)TRACK=1
PRINT FORMATS SCHOOL TO TRACK(F1.0)
VALUE LABELS SCHOOL(1)PASTEUR (2) GRANT-WHITE/
  TRACK (1)JUNE PROMOTIONS (2)FEB PROMOTIONS/
VARIABLE LABELS T1 VISUAL PERCEPTION TEST FROM SPEARMAN VPT, PART III
  T2 CUBES, SIMPLIFICATION OF BRIGHAM'S SPATIAL RELATIONS TEST
  T3 PAPER FORM BOARD--SHAPES THAT CAN BE COMBINED TO FORM A TARGET
  T4 LOZENGES FROM THORNDIKE--SHAPES FLIPPED OVER THEN IDENTIFY TARGET

  T5 GENERAL INFORMATION VERBAL TEST
  T6 PARAGRAPH COMPREHENSION TEST
  T7 SENTENCE COMPLETION TEST
  T8 WORD CLASSIFICATION--WHICH WORD NOT BELONG IN SET
  T9 WORD MEANING TEST

  T10 SPEEDED ADDITION TEST
  T11 SPEEDED CODE TEST--TRANSFORM SHAPES INTO ALPHA WITH CODE
  T12 SPEEDED COUNTING OF DOTS IN SHAPE
  T13 SPEEDED DISCRIM STRAIGHT AND CURVED CAPS

  T14 MEMORY OF TARGET WORDS
  T15 MEMORY OF TARGET NUMBERS
  T16 MEMORY OF TARGET SHAPES
  T17 MEMORY OF OBJECT-NUMBER ASSOCIATION TARGETS
```

```
   T18 MEMORY OF NUMBER-OBJECT ASSOCIATION TARGETS
   T19 MEMORY OF FIGURE-WORD ASSOCIATION TARGETS

   T20 DEDUCTIVE MATH ABILITY
   T21 MATH NUMBER PUZZLES
   T22 MATH WORD PROBLEM REASONING
   T23 COMPLETION OF A MATH NUMBER SERIES
   T24 WOODY-MCCALL MIXED MATH FUNDAMENTALS TEST
   T25 REVISION OF T3--PAPER FORM BOARD
   T26 FLAGS--POSSIBLE SUBSTITUTE FOR T4 LOZENGES
LIST VARIABLES=ALL/CASES=500/FORMAT=NUMBERED
SUBTITLE '0   DESCRIPTIVES ON ALL SUBJECTS POOLED'
FREQUENCIES VARIABLES=ID TO T26/FORMAT=NOTABLE/STATISTICS=ALL
DESCRIPTIVES VARIABLES=SCHOOL T1 TO T26
SUBTITLE '1a  SCH=PASTEUR  COMPARE WITH GORSUCH PP. 384-385'
TEMPORARY
SELECT IF (SCHOOL EQ 1)
DESCRIPTIVES VARIABLES=ALL
SUBTITLE '1b  SCH=GRANT-WH  COMPARE WITH GORSUCH PP. 384-385'
TEMPORARY
SELECT IF (SCHOOL EQ 2)
DESCRIPTIVES VARIABLES=ALL
SUBTITLE '2  COMPARE R MATRIX WITH GORSUCH P. 100'
CORRELATIONS VARIABLES=T1 TO T26
SUBTITLE '3a  REGRESSION #######    ALL CASES'
REGRESSION VARIABLES=T6 T12 T13 T23 T24/DESCRIPTIVES=MEAN STDDEV CORR/
  DEPENDENT=T6/ENTER T12 T13 T23 T24
SUBTITLE '3b  REGRESSION #######    Pasteur'
TEMPORARY
SELECT IF (SCHOOL EQ 1)
REGRESSION VARIABLES=T6 T12 T13 T23 T24/DESCRIPTIVES=MEAN STDDEV CORR/
  DEPENDENT=T6/ENTER T12 T13 T23 T24
SUBTITLE '3b  REGRESSION #######    Grant-White'
TEMPORARY
SELECT IF (SCHOOL EQ 2)
REGRESSION VARIABLES=T6 T12 T13 T23 T24/DESCRIPTIVES=MEAN STDDEV CORR/
  DEPENDENT=T6/ENTER T12 T13 T23 T24
SUBTITLE '4a  FACTOR          ALL CASES'
FACTOR VARIABLES=T1 TO T24/PRINT=DEFAULT FSCORE
SUBTITLE '4b  FACTOR          Pasteur'
TEMPORARY
SELECT IF (SCHOOL EQ 1)
FACTOR VARIABLES=T1 TO T24/PRINT=DEFAULT FSCORE/CRITERIA=FACTORS(4)
SUBTITLE '4c  FACTOR          Grant-White'
TEMPORARY
SELECT IF (SCHOOL EQ 2)
FACTOR VARIABLES=T1 TO T24/PRINT=DEFAULT FSCORE/CRITERIA=FACTORS(4)
SUBTITLE '5a  CANONICAL       ALL CASES'
MANOVA T1 TO T4 WITH T5 TO T9/PRINT=SIGNIF(EIGEN DIMENR)/
  DISCRIM(STAN COR ALPHA(.999)/DESIGN
SUBTITLE '5b  CANONICAL       Pasteur'
TEMPORARY
```

```
SELECT IF (SCHOOL EQ 1)
MANOVA T1 TO T4 WITH T5 TO T9/PRINT=SIGNIF(EIGEN DIMENR)/
  DISCRIM(STAN COR ALPHA(.999)/DESIGN
SUBTITLE '5c  CANONICAL        Grant-White'
TEMPORARY
SELECT IF (SCHOOL EQ 2)
MANOVA T1 TO T4 WITH T5 TO T9/PRINT=SIGNIF(EIGEN DIMENR)/
  DISCRIM(STAN COR ALPHA(.999)/DESIGN
subtitle '6a  z scores for Pasteur only'
temporary
select if (school eq 1)
descriptives variables=t1 (pzt1) t2 (pzt2) t3 (pzt3) t4 (pzt4) t5 (pzt5)
  t6 (pzt6) t7 (pzt7) t8 (pzt8) t9 (pzt9) t10 (pzt10) t11 (pzt11)
  t12 (pzt12) t13 (pzt13) t14 (pzt14) t15 (pzt15) t16 (pzt16)
  t17 (pzt17) t18 (pzt18) t19 (pzt19) t20 (pzt20) t21 (pzt21)
  t22 (pzt22) t23 (pzt23) t24 (pzt24)
subtitle '6b  z scores for Grant-White only'
temporary
select if (school eq 2)
descriptives variables=t1 (gzt1) t2 (gzt2) t3 (gzt3) t4 (gzt4) t5 (gzt5)
  t6 (gzt6) t7 (gzt7) t8 (gzt8) t9 (gzt9) t10 (gzt10) t11 (gzt11)
  t12 (gzt12) t13 (gzt13) t14 (gzt14) t15 (gzt15) t16 (gzt16)
  t17 (gzt17) t18 (gzt18) t19 (gzt19) t20 (gzt20) t21 (gzt21)
  t22 (gzt22) t23 (gzt23) t24 (gzt24)
subtitle '7  regression invariance analysis'
compute yhat11= (.032125*pzt12)-(.031195*pzt13)+(.383641*pzt23)+
  (.170181*pzt24)
compute yhat12=-(.252179*pzt12)+(.215562*pzt13)+(.295226*pzt23)+
  (.333045*pzt24)
compute yhat21= (.032125*gzt12)-(.031195*gzt13)+(.383641*gzt23)+
  (.170181*gzt24)
compute yhat22=-(.252179*gzt12)+(.215562*gzt13)+(.295226*gzt23)+
  (.333045*gzt24)
variable labels
  yhat11 'group 1 data  group 1 weights'
  yhat12 'group 1 data  group 2 weights'
  yhat21 'group 2 data  group 1 weights'
  yhat22 'group 2 data  group 2 weights'
correlations variables=t6 yhat11 to yhat22/statistics=descriptives
subtitle '8  factor score invariance analysis'
compute fs111= (.03613*pzt1)-(.05198*pzt2)-(.03485*pzt3)-(.09427*pzt4)
  +(.23775*pzt5)+(.22213*pzt6)+(.25641*pzt7)+(.20342*pzt8)+(.20066*pzt9)
  -(.02154*pzt10)+(.03008*pzt11)-(.08539*pzt12)-(.09797*pzt13)
  +(.01041*pzt14)-(.04524*pzt15)-(.04415*pzt16)-(.00964*pzt17)
  +(.05097*pzt18)-(.03277*pzt19)-(.02896*pzt20)-(.00314*pzt21)
  +(.08612*pzt22)+(.03258*pzt23)-(.01461*pzt24)
compute fs112= (.18800*pzt1)+(.25519*pzt2)+(.21754*pzt3)+(.25735*pzt4)
  -(.08361*pzt5)-(.04758*pzt6)-(.06830*pzt7)-(.03259*pzt8)-(.00839*pzt9)
  -(.15523*pzt10)-(.07765*pzt11)-(.01425*pzt12)+(.06293*pzt13)
  -(.03819*pzt14)+(.02857*pzt15)+(.09627*pzt16)-(.10607*pzt17)
  -(.08815*pzt18)+(.03066*pzt19)+(.20182*pzt20)+(.09159*pzt21)
  +(.12328*pzt22)+(.15888*pzt23)+(.02850*pzt24)
```

```
compute fs113=-(.05855*pzt1)-(.06732*pzt2)+(.00877*pzt3)-(.00020*pzt4)
  -(.02595*pzt5)-(.05440*pzt6)-(.07813*pzt7)-(.07813*pzt8)-(.03323*pzt9)
  +(.34434*pzt10)+(.23000*pzt11)+(.34184*pzt12)+(.27168*pzt13)
  -(.09559*pzt14)-(.11519*pzt15)+(.02875*pzt16)+(.08931*pzt17)
  -(.06895*pzt18)+(.05849*pzt19)-(.06092*pzt20)+(.13054*pzt21)
  -(.01656*pzt22)+(.01494*pzt23)+(.18016*pzt24)
compute fs114= (.00734*pzt1)-(.06730*pzt2)-(.15560*pzt3)+(.01818*pzt4)
  -(.04572*pzt5)+(.02127*pzt6)-(.01511*pzt7)+(.04886*pzt8)-(.01324*pzt9)
  -(.02159*pzt10)+(.01804*pzt11)-(.11566*pzt12)-(.07840*pzt13)
  +(.32932*pzt14)+(.31381*pzt15)+(.16406*pzt16)+(.23123*pzt17)
  +(.27537*pzt18)+(.14391*pzt19)+(.05311*pzt20)-(.02052*pzt21)
  -(.04619*pzt22)-(.00248*pzt23)+(.02203*pzt24)
compute fs211= (.03613*gzt1)-(.05198*gzt2)-(.03485*gzt3)-(.09427*gzt4)
  +(.23775*gzt5)+(.22213*gzt6)+(.25641*gzt7)+(.20342*gzt8)+(.20066*gzt9)
  -(.02154*gzt10)+(.03008*gzt11)-(.08539*gzt12)-(.09797*gzt13)
  +(.01041*gzt14)-(.04524*gzt15)-(.04415*gzt16)-(.00964*gzt17)
  +(.05097*gzt18)-(.03277*gzt19)-(.02896*gzt20)-(.00314*gzt21)
  +(.08612*gzt22)+(.03258*gzt23)-(.01461*gzt24)
compute fs212= (.18800*gzt1)+(.25519*gzt2)+(.21754*gzt3)+(.25735*gzt4)
  -(.08361*gzt5)-(.04758*gzt6)-(.06830*gzt7)-(.03259*gzt8)-(.00839*gzt9)
  -(.15523*gzt10)-(.07765*gzt11)-(.01425*gzt12)+(.06293*gzt13)
  -(.03819*gzt14)+(.02857*gzt15)+(.09627*gzt16)-(.10607*gzt17)
  -(.08815*gzt18)+(.03066*gzt19)+(.20182*gzt20)+(.09159*gzt21)
  +(.12328*gzt22)+(.15888*gzt23)+(.02850*gzt24)
compute fs213=-(.05855*gzt1)-(.06732*gzt2)+(.00877*gzt3)-(.00020*gzt4)
  -(.02595*gzt5)-(.05440*gzt6)-(.07813*gzt7)-(.07813*gzt8)-(.03323*gzt9)
  +(.34434*gzt10)+(.23000*gzt11)+(.34184*gzt12)+(.27168*gzt13)
  -(.09559*gzt14)-(.11519*gzt15)+(.02875*gzt16)+(.08931*gzt17)
  -(.06895*gzt18)+(.05849*gzt19)-(.06092*gzt20)+(.13054*gzt21)
  -(.01656*gzt22)+(.01494*gzt23)+(.18016*gzt24)
compute fs214= (.00734*gzt1)-(.06730*gzt2)-(.15560*gzt3)+(.01818*gzt4)
  -(.04572*gzt5)+(.02127*gzt6)-(.01511*gzt7)+(.04886*gzt8)-(.01324*gzt9)
  -(.02159*gzt10)+(.01804*gzt11)-(.11566*gzt12)-(.07840*gzt13)
  +(.32932*gzt14)+(.31381*gzt15)+(.16406*gzt16)+(.23123*gzt17)
  +(.27537*gzt18)+(.14391*gzt19)+(.05311*gzt20)-(.02052*gzt21)
  -(.04619*gzt22)-(.00248*gzt23)+(.02203*gzt24)
compute fs221=-(.05942*gzt1)-(.06284*gzt2)-(.03719*gzt3)-(.02332*gzt4)
  +(.23536*gzt5)+(.25056*gzt6)+(.27651*gzt7)+(.17620*gzt8)+(.26728*gzt9)
  -(.00634*gzt10)-(.03412*gzt11)-(.08810*gzt12)-(.04231*gzt13)
  -(.00362*gzt14)-(.06587*gzt15)-(.10440*gzt16)-(.02489*gzt17)
  -(.12885*gzt18)-(.01861*gzt19)+(.07507*gzt20)-(.05779*gzt21)
  +(.06915*gzt22)+(.05464*gzt23)+(.06746*gzt24)
compute fs222= (.27437*gzt1)+(.30421*gzt2)+(.20040*gzt3)+(.29744*gzt4)
  -(.03608*gzt5)-(.07273*gzt6)-(.08955*gzt7)-(.01198*gzt8)-(.06760*gzt9)
  -(.15925*gzt10)-(.06617*gzt11)+(.02109*gzt12)+(.12665*gzt13)
  -(.10132*gzt14)-(.00971*gzt15)+(.14043*gzt16)-(.17071*gzt17)
  +(.08030*gzt18)-(.03707*gzt19)+(.07865*gzt20)+(.14348*gzt21)
  +(.06203*gzt22)+(.12554*gzt23)-(.11178*gzt24)
compute fs223=-(.02782*gzt1)-(.06587*gzt2)-(.01860*gzt3)-(.07279*gzt4)
  -(.00039*gzt5)-(.06840*gzt6)-(.00928*gzt7)+(.03046*gzt8)-(.08271*gzt9)
  +(.37219*gzt10)+(.21552*gzt11)+(.33573*gzt12)+(.22256*gzt13)
  -(.07867*gzt14)-(.09831*gzt15)-(.10957*gzt16)+(.03516*gzt17)
```

```
  +(.07278*gzt18)+(.01905*gzt19)-(.09283*gzt20)+(.12063*gzt21)
  -(.06786*gzt22)-(.01729*gzt23)+(.18538*gzt24)
 compute fs224=-(.03557*gzt1)-(.08359*gzt2)-(.02880*gzt3)-(.07783*gzt4)
  -(.09318*gzt5)-(.00593*gzt6)-(.08556*gzt7)-(.07281*gzt8)-(.01840*gzt9)
  -(.05343*gzt10)+(.06830*gzt11)-(.10645*gzt12)-(.14094*gzt13)
  +(.33244*gzt14)+(.32409*gzt15)+(.24912*gzt16)+(.33842*gzt17)
  +(.17023*gzt18)+(.18221*gzt19)+(.07819*gzt20)-(.03113*gzt21)
  +(.08093*gzt22)-(.00647*gzt23)+(.03025*gzt24)
 compute fs121=-(.05942*pzt1)-(.06284*pzt2)-(.03719*pzt3)-(.02332*pzt4)
  +(.23536*pzt5)+(.25056*pzt6)+(.27651*pzt7)+(.17620*pzt8)+(.26728*pzt9)
  -(.00634*pzt10)-(.03412*pzt11)-(.08810*pzt12)-(.04231*pzt13)
  -(.00362*pzt14)-(.06587*pzt15)-(.10440*pzt16)-(.02489*pzt17)
  -(.12885*pzt18)-(.01861*pzt19)+(.07507*pzt20)-(.05779*pzt21)
  +(.06915*pzt22)+(.05464*pzt23)+(.06746*pzt24)
 compute fs122= (.27437*pzt1)+(.30421*pzt2)+(.20040*pzt3)+(.29744*pzt4)
  -(.03608*pzt5)-(.07273*pzt6)-(.08955*pzt7)-(.01198*pzt8)-(.06760*pzt9)
  -(.15925*pzt10)-(.06617*pzt11)+(.02109*pzt12)+(.12665*pzt13)
  -(.10132*pzt14)-(.00971*pzt15)+(.14043*pzt16)-(.17071*pzt17)
  +(.08030*pzt18)-(.03707*pzt19)+(.07865*pzt20)+(.14348*pzt21)
  +(.06203*pzt22)+(.12554*pzt23)-(.11178*pzt24)
 compute fs123=-(.02782*pzt1)-(.06587*pzt2)-(.01860*pzt3)-(.07279*pzt4)
  -(.00039*pzt5)-(.06840*pzt6)-(.00928*pzt7)+(.03046*pzt8)-(.08271*pzt9)
  +(.37219*pzt10)+(.21552*pzt11)+(.33573*pzt12)+(.22256*pzt13)
  -(.07867*pzt14)-(.09831*pzt15)-(.10957*pzt16)+(.03516*pzt17)
  +(.07278*pzt18)+(.01905*pzt19)-(.09283*pzt20)+(.12063*pzt21)
  -(.06786*pzt22)-(.01729*pzt23)+(.18538*pzt24)
 compute fs124=-(.03557*pzt1)-(.08359*pzt2)-(.02880*pzt3)-(.07783*pzt4)
  -(.09318*pzt5)-(.00593*pzt6)-(.08556*pzt7)-(.07281*pzt8)-(.01840*pzt9)
  -(.05343*pzt10)+(.06830*pzt11)-(.10645*pzt12)-(.14094*pzt13)
  +(.33244*pzt14)+(.32409*pzt15)+(.24912*pzt16)+(.33842*pzt17)
  +(.17023*pzt18)+(.18221*pzt19)+(.07819*pzt20)-(.03113*pzt21)
  +(.08093*pzt22)-(.00647*pzt23)+(.03025*pzt24)
 variable labels
  fs111 'grp 1 data   grp 1 weights    factor I'
  fs112 'grp 1 data   grp 1 weights    factor II'
  fs113 'grp 1 data   grp 1 weights    factor III'
  fs114 'grp 1 data   grp 1 weights    factor IV'
  fs211 'grp 2 data   grp 1 weights    factor I'
  fs212 'grp 2 data   grp 1 weights    factor II'
  fs213 'grp 2 data   grp 1 weights    factor III'
  fs214 'grp 2 data   grp 1 weights    factor IV'
  fs221 'grp 2 data   grp 2 weights    factor I'
  fs222 'grp 2 data   grp 2 weights    factor II'
  fs223 'grp 2 data   grp 2 weights    factor III'
  fs224 'grp 2 data   grp 2 weights    factor IV'
  fs121 'grp 1 data   grp 2 weights    factor I'
  fs122 'grp 1 data   grp 2 weights    factor II'
  fs123 'grp 1 data   grp 2 weights    factor III'
  fs124 'grp 1 data   grp 2 weights    factor IV'
 correlations variables=fs111 to fs124/statistics=descriptives
 subtitle '9  canonical invariance analysis'
 compute crit111= (.90048*pzt1)+(.09741*pzt2)+(.05615*pzt3)+(.08840*pzt4)
```

```
compute crit112= (.50911*pzt1)-(.68916*pzt2)+(.33707*pzt3)-(.69424*pzt4)
compute pred111=-(.31370*pzt5)+(.50778*pzt6)-(.46168*pzt7)+(.50972*pzt8)
  +(.77965*pzt9)
compute pred112=(1.16044*pzt5)+(.60716*pzt6)+(.09385*pzt7)-(.75882*pzt8)
  -(.72138*pzt9)
compute crit121= (.34770*pzt1)+(.09867*pzt2)+(.35597*pzt3)+(.52233*pzt4)
compute crit122=-(.09842*pzt1)+(1.09157*pzt2)-(.23021*pzt3)
  -(.36428*pzt4)
compute pred121= (.42959*pzt5)+(.31092*pzt6)-(.19267*pzt7)+(.51866*pzt8)
  +(.09882*pzt9)
compute pred122=(1.09361*pzt5)+(.66983*pzt6)-(.37908*pzt7)-(.89606*pzt8)
  -(.55634*pzt9)
compute crit211= (.90048*gzt1)+(.09741*gzt2)+(.05615*gzt3)+(.08840*gzt4)
compute crit212= (.50911*gzt1)-(.68916*gzt2)+(.33707*gzt3)-(.69424*gzt4)
compute pred211=-(.31370*gzt5)+(.50778*gzt6)-(.46168*gzt7)+(.50972*gzt8)
  +(.77965*gzt9)
compute pred212=(1.16044*gzt5)+(.60716*gzt6)+(.09385*gzt7)-(.75882*gzt8)
  -(.72138*gzt9)
compute crit221= (.34770*gzt1)+(.09867*gzt2)+(.35597*gzt3)+(.52233*gzt4)
compute crit222=-(.09842*gzt1)+(1.09157*gzt2)-(.23021*gzt3)
  -(.36428*gzt4)
compute pred221= (.42959*gzt5)+(.31092*gzt6)-(.19267*gzt7)+(.51866*gzt8)
  +(.09882*gzt9)
compute pred222=(1.09361*gzt5)+(.66983*gzt6)-(.37908*gzt7)-(.89606*gzt8)
  -(.55634*gzt9)
variable labels
  crit111 'criterion  grp 1 data  grp 1 weights  function 1'
  crit112 'criterion  grp 1 data  grp 1 weights  function 2'
  crit121 'criterion  grp 1 data  grp 2 weights  function 1'
  crit122 'criterion  grp 1 data  grp 2 weights  function 2'
  pred111 'predictor  grp 1 data  grp 1 weights  function 1'
  pred112 'predictor  grp 1 data  grp 1 weights  function 2'
  pred121 'predictor  grp 1 data  grp 2 weights  function 1'
  pred122 'predictor  grp 1 data  grp 2 weights  function 2'
  crit211 'criterion  grp 2 data  grp 1 weights  function 1'
  crit212 'criterion  grp 2 data  grp 1 weights  function 1'
  crit221 'criterion  grp 2 data  grp 2 weights  function 1'
  crit222 'criterion  grp 2 data  grp 2 weights  function 2'
  pred211 'predictor  grp 2 data  grp 1 weights  function 1'
  pred212 'predictor  grp 2 data  grp 1 weights  function 1'
  pred221 'predictor  grp 2 data  grp 2 weights  function 1'
  pred222 'predictor  grp 2 data  grp 2 weights  function 2'
correlations variables=crit111 to pred222/statistics=descriptives
```

<u>Note</u>. Lower case commands were inserted after an initial run was conducted to obtain the values required in the second and last run.