

Factor Indeterminacy in Generalizability Theory

David G. Ward
Fordham University

Generalizability theory and common factor analysis are based upon the random effects model of the analysis of variance, and both are subject to the factor indeterminacy problem: The unobserved random variables (common factor scores or universe scores) are indeterminate. In the one-facet (repeated measures) design, the extent to which true or universe scores and common factor scores are not uniquely defined is shown to be a function of the dependability (reliability) of the data. The minimum possible correlation between equivalent common factor scores is a lower bound estimate of reliability.

The equivalence of the random effects model of the analysis of variance and the model of factor analysis is well-established (Bock, 1960; Burt, 1947; Creasy, 1957), as is the equivalence of the classical test theory model (the true score model) and the common factor model (e.g., Steiger & Schönemann, 1978). Generalizability theory (Cronbach, Gleser, Nanda, & Rajaratnam, 1972) extends classical test theory by ascribing variations in observations to specific sources (facets or conditions of measurement). Because generalizability theory is based upon the random effects model, problems associated with the factor model may also be associated with generalizability studies.

One such problem in the factor model is the lack of determinacy of the factor scores which presum-

ably underlie the observed scores (e.g., Cattell, 1973, p. 303; Guttman, 1955; Schönemann, 1971). In short, there are infinitely many solutions for the unobserved factor scores, and thus no unique solution for the factor model as defined in the population. Universe scores (person variables) are unobserved common factor scores, and they too are subject to the indeterminacy problem.

There is considerable controversy surrounding the factor indeterminacy issue (e.g., Jensen, 1983; McDonald, 1977; Schönemann, 1983; Steiger & Schönemann, 1978). For example, McDonald (1974, 1977) presented a case for factor variables as being unique but unknown. The present approach takes a strictly mathematical view of the linear model used in generalizability theory, namely that a system of linear equations is indeterminate if there are more unknown variables (common and unique factors) than known (observed) variables. This study examined the degree of factor indeterminacy in the one-facet (repeated measures) design used in generalizability theory.

Generalizability studies focus on estimating the variance components associated with sources of variation rather than on estimating individual universe scores or the effects associated with the measurement conditions, as in decision studies (e.g., Cronbach et al., 1972, pp. 16–17). Variability in scores is examined rather than estimation of universe scores, because the levels (e.g., observers, situations, items) are assumed to be randomly selected from some relevant universe, as in random

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 10, No. 2, June 1986, pp. 159–165
© Copyright 1986 Applied Psychological Measurement Inc.
0146-6216/86/020159-07\$1.60

effects analysis of variance. There is considerable literature on estimation of true scores (e.g., Lord & Novick, 1968, pp. 152–153) and universe scores (e.g., Cronbach et al., 1972, pp. 73ff.), but the factor indeterminacy problem is not simply one of estimating universe scores or person effects. Rather, there is no way to determine a person's universe score uniquely, even if a person's universe score is defined a priori as a parameter—that is, the expectation of observed scores over all conditions of measurement (randomly parallel tests), as in true score theory. This is equivalent to taking the expectation over all possible levels of the facet, which is not feasible, especially with large universes such as observers, occasions, or situations.

It will be shown that both the generalizability coefficient and an index of factor indeterminacy are estimates of reliability. The lowest possible reliability is the minimum correlation between equivalent universe (or common factor) scores. While the goal of this study was determining the number of observers, occasions, or situations needed to obtain acceptable reliability in direct behavioral observation or rating studies, there are also implications for factor analytic test construction (with respect to choosing the number of items in a test). The pragmatic question, then, is: To what extent, if any, does this indeterminacy have practical significance? What is the number of items or observations needed to achieve a given level of reliable measurement?

Overview of Generalizability Theory

Attention here is limited to fully crossed (balanced) one-facet (repeated-measures) designs, though the results may extend directly to other models. Working with deviation scores allows considerable simplification without any loss in generality, because the variance component estimates are unaffected by subtracting the population mean from each observation. Thus, an observation in deviation score form may be decomposed into independent random variables corresponding to the sources of variation in the experimental design, plus a random error variable.

A typical application of a one-facet design, such as $P \times S$ (person \times situation), is collecting behavioral observations or ratings on n people in n_s different situations. The n people are assumed to be randomly sampled from the population of all such people, and the n_s levels of facet S are assumed to be randomly selected from the universe of all such conditions of measurement. A sample deviation score (Cronbach et al., 1972, p. 26) is represented as

$$y_{ij} = p_i + s_j + ps_{ij} + e_{ij} \quad (1)$$

where y_{ij} is the observed deviation score for person i under condition j ;

p_i is the effect of person i (the universe score for person i minus the population mean);

s_j is the effect of condition j ;

ps_{ij} is the $P \times S$ interaction effect of person i under condition j ; and

e_{ij} is random error.

Because differences in means for levels of facet S do not affect the person-related variance component estimates, the different effects due to facet S can be ignored for the present purposes. The model assumptions are:

$$E(y) = E(p) = E(ps) = E(e) = 0 \quad (2)$$

$$\text{Var}(y) = \sigma_y^2 \quad (3)$$

$$\text{Var}(p) = \sigma_p^2 \quad (4)$$

$$\text{Var}(ps) = \sigma_{ps}^2 \quad (5)$$

$$\text{Var}(e) = \sigma_e^2 \quad (6)$$

and p , ps , and e are independent. Hence,

$$\sigma_y^2 = \sigma_p^2 + \sigma_{ps}^2 + \sigma_e^2 \quad (7)$$

In the usual sample estimation problem, ps_{ij} and e_{ij} cannot be measured separately and are confounded (the residual pse_{ij}). Thus, the model actually used is

$$y_{ij} = p_i + pse_{ij} \quad (8)$$

where p_i and pse_{ij} are independent, so that

$$\sigma_y^2 = \sigma_p^2 + \sigma_{pse}^2 \quad (9)$$

The variance components may be estimated using the Cornfield-Tukey algorithm,

$$\hat{\sigma}_p^2 = (MS_p - MS_{pse})/n_s \quad (10)$$

and

$$\hat{\sigma}_{pse}^2 = MS_{pse} \quad (11)$$

where MS_p and MS_{pse} are the mean squares for persons and residual, respectively.

Considering the population case, the so-called unit-sample generalizability coefficient (g-coefficient) ρ_i^2 is an intraclass correlation coefficient which measures the reliability of a single judge, or a single item in a test, as in traditional reliability (e.g., McKeon, 1969; Winer, 1971, pp. 283–286):

$$\rho_i^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{pse}^2} \quad (12)$$

The generalizability of the mean of n_s measures, the full-sample generalizability coefficient (or coefficient alpha; Mulaik, 1972, p. 209), is given by

$$\rho_{n_s}^2 = \frac{\sigma_p^2}{\sigma_p^2 + (1/n_s)\sigma_{pse}^2} \quad (13)$$

The full-sample g-coefficient corresponds to the reliability of a whole test comprised of n_s items, which may also be computed using the Spearman-Brown prophecy formula.

Schönemann (1971) has shown that for a given set of observations which satisfy the factor model, a minimally correlated equivalent set of uncorrelated factor scores may be generated which correspond to the same observations. This minimally correlated equivalent set of orthogonal factors represents the worst possible case of factor indeterminacy, and such results may be used to evaluate the random effects model as used in generalizability theory.

Factor Indeterminacy

To phrase the random effects model in factor analysis terms, consider the model in Equation 8. This discussion holds for any number of repeated measures, but for illustration, consider the case $n_s = 2$. This discussion is based on Schönemann (1971, 1983) and Steiger and Schönemann (1978). The model in Equation 8 may be written

$$y_{ij} = p_i + pse_{ij} = \sigma_p x_i + \sigma_{pse} z_{ij} \quad (14)$$

where $E(x) = E(z_j) = 0$;

$$\text{Var}(x) = \text{Var}(z_j) = 1; \text{ and}$$

x and z_j are mutually independent ($j = 1, 2$).

The two observed random variables y_j in $\mathbf{y}' = (y_1, y_2)$ may be written in vector notation as

$$\mathbf{y} = (\mathbf{a}, \mathbf{U}) \begin{pmatrix} x \\ \mathbf{z} \end{pmatrix} = \mathbf{a}x + \mathbf{U}\mathbf{z} \quad (15)$$

where x is an unobserved random variable, the standardized “common factor”;

the z_j in $\mathbf{z}' = (z_1, z_2)$ are two unobserved random variables, standardized “unique factors”;

\mathbf{a} is a 2×1 vector of factor loadings, the common factor pattern, where $\mathbf{a}' = (\sigma_p, \sigma_p)$;

\mathbf{U} is a positive definite diagonal matrix, the unique factor pattern, where $\mathbf{U} = \sigma_{pse} \mathbf{I}_2$; and

the matrix $(\mathbf{a}, \mathbf{U}) = (\sigma_p, \sigma_{pse} \mathbf{I})$ is the total pattern (cf. Schönemann, 1971, p. 22).

The “person variable” is $p = \sigma_p x$, and the person \times situation residual variables are $pse_1 = \sigma_{pse} z_1$ and $pse_2 = \sigma_{pse} z_2$. Thus, the one-facet ($P \times S$) design corresponds to the two-factor model, where σ_p^2 is the person variance component (common variance or communality), σ_{pse}^2 is the residual variance component (unique variance, consisting of “true” specific and error variance), and the covariance matrix of the observations is

$$\begin{aligned} \Sigma &= \sigma_p \sigma_p' + \sigma_{pse}^2 \mathbf{I} \\ &= \begin{bmatrix} \sigma_p^2 + \sigma_{pse}^2 & \sigma_p^2 \\ \sigma_p^2 & \sigma_p^2 + \sigma_{pse}^2 \end{bmatrix} \quad (16) \end{aligned}$$

The universe (or true) scores correspond to the first centroid of the observed data. (The possibility of the person variable being composed of more than one common factor is not pursued here, nor are the various other generalizability models.)

In the usual sample estimation problem,

$$\mathbf{Y} = \hat{\sigma}_p \mathbf{x}' + \hat{\sigma}_{pse} \mathbf{I} \mathbf{Z} \quad (17)$$

where $\mathbf{Y}(2 \times n)$ are the observed deviation scores; $\mathbf{x}'(1 \times n)$ is a vector of standardized “common factor scores”;

$\mathbf{Z}(2 \times n)$ is a matrix of standardized and uncorrelated “unique factor scores”; and

$\hat{\sigma}_p$ and $\hat{\sigma}_{pse}$ are sample estimates of σ_p and σ_{pse} .
 Then,

$$\hat{\Sigma} = \mathbf{Y}\mathbf{Y}'/(n-1) \quad (18)$$

The essence of factor indeterminacy is that an infinite number of solutions for x and z exist which satisfy Equation 15. More formally, Schönemann (1971) has shown that for any set of factors (x, z') which satisfy Equation 15, another set of equivalent factors

$$\begin{pmatrix} x^* \\ z^* \end{pmatrix} = \mathbf{T}' \begin{pmatrix} x \\ z \end{pmatrix} \quad (19)$$

may be found which also satisfy

$$\begin{aligned} (\mathbf{a}, \mathbf{U}) \begin{pmatrix} x \\ z \end{pmatrix} &= (\mathbf{a}, \mathbf{U}) \mathbf{T} \mathbf{T}' \begin{pmatrix} x \\ z \end{pmatrix} \\ &= (\mathbf{a}, \mathbf{U}) \mathbf{T}' \begin{pmatrix} x \\ z \end{pmatrix} \\ &= (\mathbf{a}, \mathbf{U}) \begin{pmatrix} x^* \\ z^* \end{pmatrix} \end{aligned} \quad (20)$$

and Equation 15, where the transformation matrix \mathbf{T} is an orthogonal right unit of (\mathbf{a}, \mathbf{U}) (Schönemann, 1971, p. 23). An orthogonal right unit of the matrix (\mathbf{a}, \mathbf{U}) has the properties: $(\mathbf{a}, \mathbf{U})\mathbf{T} = (\mathbf{a}, \mathbf{U})$ and $\mathbf{T}\mathbf{T}' = \mathbf{T}'\mathbf{T} = \mathbf{I}$. \mathbf{T} is also the correlation matrix between equivalent sets of factors:

$$\text{cov} \left[\begin{pmatrix} x \\ z \end{pmatrix}, \begin{pmatrix} x^* \\ z^* \end{pmatrix} \right] = \left[\text{var} \begin{pmatrix} x \\ z \end{pmatrix} \right] \mathbf{T} = \mathbf{T} \quad (21)$$

The \mathbf{T} which minimizes the correlation between factor pairs is given by

$$\mathbf{T}_{\min} = 2 \begin{pmatrix} \mathbf{a}' \\ \mathbf{U}' \end{pmatrix} \Sigma^{-1} (\mathbf{a}, \mathbf{U}) - \mathbf{I}_{n_s+1} \quad (22)$$

(Schönemann, 1971, Equation 5.2, based on Guttman, 1955, and Heermann, 1966). Although \mathbf{T}_{\min} minimizes the sum of the correlations between equivalent factor pairs [i.e., $\text{tr}(\mathbf{T}_{\min})$], which includes both common and unique factors, the minimum possible correlation between the equivalent common factor scores may even be lower (cf. Schönemann & Wang, 1972). In general, the minimum average correlation between equivalent sets

of uncorrelated factors is given by

$$\begin{aligned} \tau &= [1/(p+m)] \text{tr}(\mathbf{T}_{\min}) \\ &= [1/(p+m)] \sum_{k=1}^{p+m} t_{kk} \\ &= (p-m)/(p+m) \end{aligned} \quad (23)$$

where p is the number of observed random variables,

m is the number of common factors, and t_{kk} are the diagonal elements of \mathbf{T}_{\min} (i.e., the minimum correlations between equivalent sets of common and unique factors; Schönemann, 1971).

In a one-facet generalizability model, then, $\tau = (n_s - 1)/(n_s + 1)$, and for $n_s = 2$, $\tau = 1/3$.

Perhaps more relevant is the correlation between equivalent common factors. Schönemann (1971, p. 27) discussed the minimum average correlation between equivalent sets of m uncorrelated common factors, given by

$$\begin{aligned} \tau_x &= (1/m) \sum_{j=1}^m t_{jj} \\ &= (1/m) \text{tr}(2\mathbf{A}'\Sigma^{-1}\mathbf{A} - \mathbf{I}_m) \end{aligned} \quad (24)$$

where τ_x is the mean of the first m diagonal elements of \mathbf{T}_{\min} . For the one-common factor model used in generalizability theory, this minimum correlation between equivalent sets of universe scores based on minimizing $\text{tr}(\mathbf{T}_{\min})$ is t_{11} , element (1,1) of \mathbf{T}_{\min} . For convenience, we can define this correlation as r_{\min} , which is given by

$$r_{\min} = t_{11} = 2\mathbf{a}'\Sigma^{-1}\mathbf{a} - 1 \quad (25)$$

(from Schönemann, 1971, p. 27, and from Equation 22). Substituting σ_p and σ_{pse} into this definition of r_{\min} yields the following in terms of the variance components:

$$\begin{aligned} r_{\min} &= \frac{\sigma_p^2 - (1/n_s)\sigma_{pse}^2}{\sigma_p^2 + (1/n_s)\sigma_{pse}^2} \\ &= \rho_{n_s}^2 - \frac{(1/n_s)\sigma_{pse}^2}{\sigma_p^2 + (1/n_s)\sigma_{pse}^2} \end{aligned} \quad (26)$$

The minimum correlation between equivalent common factors is thus the full-sample generalizability minus the ratio of the residual variance to the observed score variance used in the denominator of the full-sample g -coefficient.

r_{\min} reflects the degree of variability in possible values for the universe scores of a given person in the population, as a lower bound estimate of reliability. r_{\min} is the cosine of the angle formed by possible equivalent universe score vectors: The smaller the angle, the less the indeterminacy.

The ratio of r_{\min} to $\rho_{n_s}^2$ is

$$\frac{r_{\min}}{\rho_{n_s}^2} = \frac{\sigma_p^2 - (1/n_s) \sigma_{pse}^2}{\sigma_p^2} = 1 - \frac{\sigma_{pse}^2}{n_s \sigma_p^2} \quad (27)$$

For two standardized observed variables ($n_s = 2$), the following relationships hold (where r_{12} is the Pearson product-moment correlation between the two measures):

$$\rho_1^2 = r_{\min} \quad (28)$$

and

$$r_{\min} = \frac{3r_{12} - 1}{r_{12} + 1} \quad (29)$$

Equation 29 is derived by substituting Equation 28 in Equation 12, solving for

$$\sigma_p^2 = \sigma_{pse}^2 r_{12} / (1 - r_{12}) \quad (30)$$

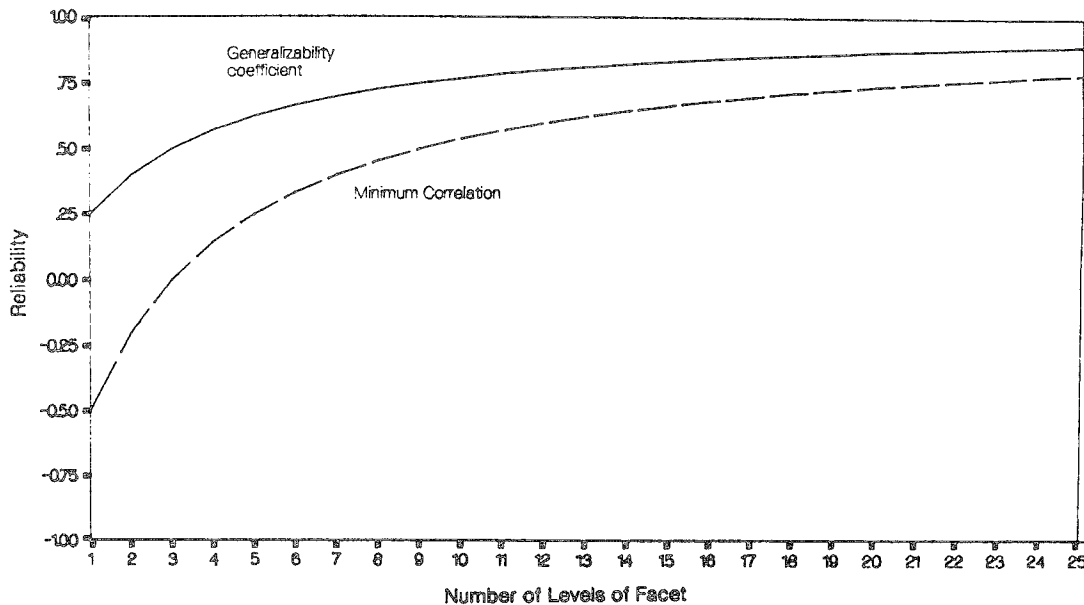
and substituting this result in Equation 26. Thus, if $r_{12} = 1$, then $\rho_1^2 = 1$ and $r_{\min} = 1$; and if $r_{12} = 0$, then $\rho_1^2 = 0$ and $r_{\min} = -1$. If the observed PPM correlation is as low as $r_{12} = 1/3$, then the correlation between equivalent common factors (universe scores) may be as low as $r_{\min} = 0$.

For studies involving more than $n_s = 2$ (i.e., more than 2 observers or items), the relationship between reliability ($\rho_{n_s}^2$ and r_{\min}) and n_s is shown in the accompanying figure for a representative value of the unit-sample g-coefficient ($\rho_1^2 = .25$, based on the relationship between σ_p^2 and σ_{pse}^2). The full-sample generalizability coefficient, which is the reliability of an n_s -item extended test and represents the Spearman-Brown prophecy graph, and the minimum correlation between equivalent universe scores, are both monotonically increasing functions of the number of items or levels of the facet. The difference between the g-coefficient (top curve) and r_{\min} (lower curve) is

$$\sigma_{pse}^2 / (n_s \sigma_p^2 + \sigma_{pse}^2) \quad (31)$$

as in Equation 26. Other initial values of the unit-

Figure 1
 Generalizability and Minimum Correlation Between Equivalent Common Factors
 as a Function of Number of Levels of the Facet



sample g -coefficient yield similar sorts of curves. For different measurement problems, there may be a certain "optimal" number of items, beyond which further improvement in reliability is relatively less cost-effective in terms of time and resources expended.

Discussion

On the basis of the results obtained in this study, the following conclusions are offered. In a one-facet design, the degree of indeterminacy decreases as the reliability or generalizability of the measurement procedure (i.e., instrument, observer, etc.) increases. The extent of indeterminacy (or conversely, the dependability of the data) is indexed by either the g -coefficient or the correlation between equivalent common factors from minimally correlated equivalent sets. Each level (e.g., observer, judge) may be regarded as a test or measuring instrument. According to Mulaik (1972): "In a universe of tests common-factor analysis, image analysis, canonical factor analysis, and image factor analysis all converge to the model of classic reliability theory which partitions variables into true-score and error-score parts" (p. 215). The present analysis also demonstrates the equivalence of the simplest generalizability model to the above models, while more complex and higher-order designs represent modified factor models.

Further, A. J. Conger's (personal communication, 1983) notion of the interchangeability of measurement instruments is addressed by the generalizability across levels of a facet. Both the generalizability coefficient and the correlation between equivalent common factors assess the degree of agreement in rank ordering of persons by "equivalent" or interchangeable measurement devices, such as observers. The degree of similar rank ordering is indexed by the person \times facet interaction variance component: The greater the agreement across levels, the smaller the interaction component. Thus, the generalizability increases and the indeterminacy decreases as the relative contribution of the person variance component increases and as the number of levels (n) increases (cf. Mulaik & McDonald, 1978).

This paper has dealt with a relatively simple measurement problem, namely a unidimensional scale or unitary construct. More complex measurement situations, such as item selection using domain sampling methods and construction of multi-attribute scales, demand different approaches. For example, especially in questionnaire and test construction, Cattell's (1973) method of progressive rectification using parcels of items and Comrey's (1973) factored homogeneous item dimensions are useful conceptual and methodological tools, especially in an ongoing research program. As a general rule, factor analytically constructed scales should not have fewer than six to eight items per scale, and more would be preferred to assure adequate reliability. A related issue is the problem of overfactoring, extracting too many factors or components, especially as in confirmatory factor analysis to obtain better fit (e.g., Steiger & Schönemann, 1978, p. 164). Such capitalization on chance leads to factors being operationalized by only a few items or variables and thereby being less replicable (less determinate). Generally the first principal factors or components are most robust (replicable).

Further issues currently being investigated include: relationships among the measures of factor indeterminacy discussed here, Guttman's (1955) measure based on the regression of true scores on observed scores, and a shrunken estimate based on Wherry's shrinkage formula; the degree of indeterminacy in higher-order designs; and the multivariate case, in terms of set correlation (Cohen & Cohen, 1983, pp. 487-518; Cohen & Nee, 1984).

References

- Bock, R. D. (1960). Components of variance analysis as a structural and discriminant analysis for psychological tests. *British Journal of Statistical Psychology*, *13*, 151-163.
- Burt, C. (1947). A comparison of factor analysis and analysis of variance. *British Journal of Psychology, Statistical Section*, *1*, 3-26.
- Cattell, R. B. (1973). *Personality and mood by questionnaire: A handbook of interpretive theory, psychometrics, and practical procedures*. San Francisco: Jossey-Bass.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sci-*

- ences (2nd ed.). Hillsdale NJ: Lawrence Erlbaum.
- Cohen, J., & Nee, J. C. M. (1984). Estimators for two measures of association for set correlation. *Educational and Psychological Measurement*, 44, 907–917.
- Comrey, A. L. (1973). *A first course in factor analysis*. New York: Academic Press.
- Creasy, M. A. (1957). Analysis of variance as an alternative to factor analysis. *Journal of the Royal Statistical Society, Series B (Methodological)*, 19, 318–325.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements*. New York: Wiley.
- Guttman, L. (1955). The indeterminacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, 8, 65–82.
- Heermann, E. F. (1966). The algebra of factorial indeterminacy. *Psychometrika*, 31, 539–543.
- Jensen, A. R. (1983). The definition of intelligence and factor-score indeterminacy. *The Behavioral and Brain Sciences*, 6, 313–315.
- Lord, F., & Novick, M. (1968). *Statistical theories of mental test scores*. Reading MA: Addison-Wesley.
- McDonald, R. P. (1974). The measurement of factor indeterminacy. *Psychometrika*, 39, 203–222.
- McDonald, R. P. (1977). The indeterminacy of components and the definition of common factors. *British Journal of Statistical Psychology*, 30, 165–176.
- McKeon, F. M. (1969). Relations among canonical reliability, canonical factor analysis, and multivariate association. *Psychometrika Monograph Supplement*, 34 (No. 15).
- Mulaik, S. A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.
- Mulaik, S. A., & McDonald, R. P. (1978). The effect of additional variables on factor indeterminacy in models with a single common factor. *Psychometrika*, 43, 177–192.
- Schönemann, P. H. (1971). The minimum average correlation between equivalent sets of uncorrelated factors. *Psychometrika*, 36, 21–30.
- Schönemann, P. H. (1983). Do IQ tests really measure intelligence? *The Behavioral and Brain Sciences*, 6, 311–313.
- Schönemann, P. H., & Wang, M. M. (1972). Some new results on factor indeterminacy. *Psychometrika*, 37, 61–91.
- Steiger, J. H., & Schönemann, P. H. (1978). A history of factor indeterminacy. In S. Shye (Ed.), *Theory construction and data analysis in the social sciences* (pp. 136–178). San Francisco: Jossey-Bass.
- Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.

Acknowledgments

Appreciation is expressed to Peter H. Schönemann for pointing out the factor indeterminacy problem, and to unknown reviewers for helpful comments.

Author's Address

Send requests for reprints or further information to David G. Ward, Department of Psychology, Fordham University, Bronx NY 10458.