PSYCHOMETRIKA—VOL. 53, NO. 1, 63–70
MARCH, 1988

# INTERNAL CONSISTENCY OF TESTS: ANALYSES OLD AND NEW

LEE J. CRONBACH

STANFORD UNIVERSITY

A coefficient derived from communalities of test parts has been proposed as greatest lower bound to Guttman's "immediate retest reliability." The communalities have at times been calculated from covariances between item *sets*, which tends to underestimate appreciably. When items are experimentally independent, a consistent estimate of the greatest defensible internal-consistency coefficient is obtained by factoring item covariances. In samples of modest size, this analysis capitalizes on chance; an estimate subject to less upward bias is suggested. For estimating alternate-forms reliability, communality-based coefficients are less appropriate than stratified alpha.

Key words: communality, internal-consistency, reliability.

This paper originated in a paradox. Bentler and Woodward (1980, 1983; hereafter BW) reasoned that a certain internal-consistency analysis promises "the greatest lower bound to reliability." Their illustrative coefficients, however, were strangely low. For Comrey's measure of social conformity, BW reached a coefficient of .85 whereas Comrey (1970) had obtained a coefficient of .94 from much the same cases. For the Test of English as a Foreign Language (TOEFL), BW reached .92 as "greatest lower bound"; the same data led Lord and Novick (LN, 1968, p. 91) to a coefficient of .95. All these coefficients reflect the internal consistency among parts of the test. The paradox is resolved by recognizing that these BW analyses took *sets* of items as the parts, whereas Comrey and LN analyzed unaggregated items.

What parts should be analyzed has not been specified by BW or in the pertinent papers by other authors. It will be argued here that the BW procedure should have been applied to covariances of Comrey and TOEFL items. In these instances, incorrect choice of unit of analysis undercut an otherwise brilliant technical development. This paper extends the interpretation of the BW procedure and suggests an alternative that capitalizes on chance to a smaller degree. It is pointed out also that even in the population the BW coefficient, calculated from items, tends to overestimate alternate-forms reliability.

Table 1 gives a reference list of nearly all coefficients to be discussed. The unobservable coefficients $\rho_{tt}$ and $\rho_s$ are the targets that various procedures were devised to estimate. An estimator $\rho_+$ which employs communalities of test parts was proposed by BW and others; this paper emphasizes its special case $\rho_{++}$. These were developed as estimators of $\rho_{tt}$. The more familiar alpha coefficients, based on average covariances of test parts, were developed as estimates of $\rho_s$. Tables 2 and 3 will illustrate many of the coefficients with data from Comrey's test.

I thank Edward Haertel for comments and suggestions, and Andrew Comrey for data.

Requests for reprints should be sent to Lee J. Cronbach, Center for the Study of Youth Development, Building 460, Stanford University, CA 94305.

0033-3123/88/0300-9027$00.75/0

TABLE 1

Coefficients and Their Symbols

| | |
|---|---|
| **Target reliabilities** | |
| $\rho_{tt}$ | Correlation between independent administrations of a fixed collection of items |
| $\rho_s$ | Intraclass correlation across test forms constructed by plan s with fixed strata |
| **Internal-consistency estimators** | |
| $\rho_+$ | Coefficient obtained by estimating communalities from covariances of parts of a test |
| $\rho_{++}$ | $\rho_+$ obtained from covariances of all items |
| $\rho^*$ | Coefficient obtained by estimating communalities from covariances within sets of items |
| $\alpha_s$ | Coefficient obtained using average covariances within sets of items |
| $\alpha$ | Limiting case of $\alpha_s$ where all items are treated as a single set |

## Theoretical Background

First, a brief review. Theoretical statements will rest on three assumptions: when tested twice, persons do not change from trial to trial; the sample of persons is indefinitely large; and items within a trial are experimentally independent. (That is to say, error scores are uncorrelated, where error is defined classically as the difference between the person's observed score and true score on an item.)

The correlation $\rho_{tt}$ between independent administrations of a fixed item-set is what BW sought to bound. This was called "immediate retest reliability" by Guttman (1945) and "hypothetical self-correlation" by Cronbach (1947); in this conception the true score on the test includes item-specific factors and all item-common factors. We shall be concerned also with estimates of the alternate-forms coefficient $\rho_s$, which is the ratio of true-score variance to expected observed-score variance of test forms constructed by sampling items from a domain in accord with fixed specifications $s$. In this conception the true score on the family of tests includes only the first centroid factor of the test covariances.

The coefficient $\rho_s$ applies to the family, whereas $\rho_{tt}$ describes a fixed test. Unless additional assumptions are made, values of $\rho_{tt}$ differ from test to test within a family. Where test $t$ belongs to the family defined by $s$, $\rho_s \leq E\rho_{tt}$. Here and elsewhere, the expectation is over tests within the family. ($E\rho^2_{\mu t}$, the expected squared correlation of the test score with the true score for the family, is not considered here. $\rho_s \leq E\rho^2_{\mu t}$, but for most tests, especially those constructed with a specified distribution of content, the difference is small; see Cronbach, Ikeda, & Avner, 1964; and Rajaratnam, Cronbach, & Gleser, 1965.)

It is not trivial to ask which coefficient an investigator should seek, because a lower bound to $\rho_{tt}$ may overestimate $\rho_s$. The choice should depend on the interpretation given the test. If factors orthogonal to the first centroid are nuisance variables, $\rho_s$ is more relevant than $\rho_{tt}$. To the extent that common factors beyond the first carry wanted information, $\rho_{tt}$ is more pertinent. The items of TOEFL, for example, are of no interest in themselves; in lengthening the instrument or making a new form one would use fresh items that conform to the same specifications and so as a set are targeted on the same centroid.

When subjects have taken only one test form, analysis of scores from parts—items, or sets of items—provides a basis for estimating $\rho_{tt}$ and $\rho_s$. Every internal-consistency formula can be expressed as a ratio: the sum of elements in a matrix divided by the variance of test scores. Off-diagonal cells of the matrix contain covariances of test parts; different formulas in effect embody different rules for filling diagonal cells of the numerator. And, if variances of parts are placed in the diagonal, the sum of elements equals the score variance for the denominator.

When test items have been grouped on some rational basis, $\alpha_s$ ("stratified alpha") estimates $\rho_s$ for the family of tests constructed on that basis. This type of formula originated with R. Jackson and Ferguson (1941) and evolved through papers by Lord (1956), Tryon (1957), and Rajaratnam et al. (1965), among others. The LN coefficient for TOEFL is an $\alpha_s$, and Comrey's is a matched-halves coefficient nearly equivalent to $\alpha_s$. We may think of a large domain of admissible items, and of a rule $s$ that specifies subdomains together with the respective numbers of items to be drawn from them. (In what follows, each number is assumed to be two or greater. The item sets need not be presented as formal subtests.) This sampling model is an analog of the use of a table of specifications in test construction, where the domain of items exists only in principle (Lord & Novick, 1968, pp. 234–235).

When every diagonal cell $i$, $i$ in the item-covariance matrix is filled with the average covariance of $i$ with other items in its set, the sum of all elements is the numerator for $\alpha_s$. Equivalently, in the matrix of covariances between sets, the diagonal cell for a set of $m$ items can be filled with $m^2$ times the average interitem covariance within the set. With many ways to partition the domain, there are many possible coefficients $\rho_s$. Whatever the partitioning rule, the numerator for $\alpha_s$ is an unbiased estimate of the covariance of the test with the true score corresponding to rule $s$. Moreover, the expected value of that covariance is the true-score variance, and $\alpha_s$ approximates $\rho_s$ (Cronbach, Schönemann, & McKie, 1965).

The line of thought that seeks a bound for $\rho_{tt}$ originated with Guttman (1945) and continued through papers by Bentler (1972), P. Jackson and Agunwamba (1977), Woodhouse and P. Jackson (1977), Della Riccia and Shapiro (1980), Bentler and Woodward (1980, 1983), and ten Berge, Snijders, and Zegers (1981), among others. To obtain $\rho_{tt}$ one would need to place the non-error variance for item $i$ in the $i$, $i$ cell of the covariance matrix; but the analyst cannot separate true item-specific variance from error. The recent papers in this tradition are concerned with coefficients obtained by placing estimated communalities of parts in the diagonal cells of the numerator matrix. Constrained minimum trace factor analysis (CMTFA) has been accepted for evaluating the communalities. The algorithms for CMTFA recommended by ten Berge, et al. and by Bentler and Woodward (1983) are effectively the same.

## Levels of Partitioning

Bentler and Woodward (1983, p. 237) spoke of the test as "a composite formed from unit weighted component parts", and then presented a theory of decomposition valid for any matrix of part covariances. The impression left by this and other descriptions of such methods is that a file of scored test responses yields just one coefficient $\rho_+$. The coefficient changes with the partitioning, however.

The choice among partitionings was discussed in writings on $\alpha_s$, and those ideas will help in understanding $\rho_+$. Subdomains used in test specification may be broad or narrow, indistinguishable or distinctly different. At one extreme, placing all items in one stratum produces "randomly parallel" tests, and $\alpha_s$ degenerates to $\alpha$. At the other extreme, the constructor of an alternate form may attempt to write a counterpart for each original item, to present almost exactly the same question again, at a similar level of difficulty ("matched" tests; Lord, 1955). Comrey specified his measure of conformity by listing five subdimensions such as "need for approval", and then—to neutralize the acquiescence factor—requiring two positively keyed and two negatively keyed items for each. Narrow strata can be clustered into broader ones. In Comrey's plan, a stratum is defined by both content and keying; these strata can be grouped in content strata (or, if one prefers, into direction-of-keying strata).

Finer stratification is intended to increase $\rho_s$, and it does so when in the domain the overall mean of item covariances within the narrow sets is greater than the overall mean within whatever broader set(s) the plan subdivides. When this condition holds, $\alpha_s$ is likely to increase with finer partitioning in the analysis (but not necessarily in every sample of items). Finer partitioning is likely to increase $\rho_+$ also, since partitioning moves covariances originally aggregated within the variances of broad sets out into the base for communality estimates. The mean-covariance condition stated above is neither a necessary nor a sufficient condition for increase in $\rho_+$; it is doubtful that a general condition for the increase can be stated.

## Interpretations of $\rho_+$

Communalities from the item-covariance matrix yield a $\rho_+$ that is the greatest defensible internal-consistency coefficient when the assumptions stated early in this paper are satisfied. The symbol $\rho_{++}$ will distinguish this from a $\rho_+$ derived from between-set covariances, which is expected to be smaller than $\rho_{++}$ and cannot be greater.

Also, $\alpha_s$ cannot be greater than $\rho_{++}$. This follows from an argument of Bentler and Woodward (1980, p. 254). Their coefficient $\rho_1$, calculated from a matrix with just one column of unit weights for items (the signs of the weights being determined by CMTFA), is the maximum intraclass split-half coefficient for the test. Moreover, it is not less than a $\rho_+$ from a matrix of weights having greater column rank; hence $\rho_1 \leq \rho_{++}$. Rational pairing of items to form matched halves gives the largest $\alpha_s$. Since that cannot exceed the empirical maximum $\rho_1$, $\alpha_s \leq \rho_{++}$.

It is quite possible to obtain a $\rho_+$ smaller than the $\alpha_s$ from the same item-sets. The between-set covariances from which $\rho_+$ comes count only factors that two sets share; in contrast, the diagonal cells for the numerator of $\alpha_s$ reflect only factors shared by items *within* the set. Any attempt to define homogeneous strata should create comparatively strong within-set factors.

The Bentler and BW papers offer several illustrative calculations of "greatest lower bounds"; half of them take item-sets as parts, hence are not $\rho_{++}$. Typically in these instances, published correlations or covariances of sets were the starting point, item covariances not being at hand. Even where raw data were available for a Wechsler Scale example, Bentler and Woodward (1983, p. 251) factored subtest covariances (perhaps to reduce computing time and cost). Woodhouse and P. Jackson speak of factoring "subtests" in their numerical example also; these parts were actually separate brief examinations that it seemed impractical to divide, especially as the student was allowed to choose which questions to write on (P. Jackson, personal communication, January 9, 1987).

## An Alternative Procedure

Analysis of bias in CMTFA has only begun (Zegers & Knol, 1980). Although in the population $\rho_{++}$ is a lower bound to $\rho_{tt}$, variation in covariances due to sampling of persons will tend to inflate an estimate $\hat{\rho}_{++}$. Bentler and Woodward (1983, pp. 246–249) discuss this and other statistical questions.

With samples of modest size, one would expect much less bias in the sum of communalities from several matrices of low order than in the sum of communalities from a single large matrix, because with low order fewer weights are fitted. This suggests that applying CMTFA to smaller covariance matrices can reduce upward bias (and reduce computing cost). To obtain what I shall call $\rho^*$, each set of items would be analyzed in turn, communalities for items in part P being evaluated by CMTFA of their intercovariances. Filling the diagonal of the matrix for all items with such communalities, and then summing elements, gives the numerator for $\rho^*$. (Bentler and Woodward, 1983, p. 243, employ the symbol $\rho^*$ for a peripheral concept unrelated to this coefficient.)

By way of a limited numerical demonstration, a 20-item test was defined with five strata, having population item variances equal to 1.00 and item covariances equal to .25 within strata and .15 across strata, that is, there is a general factor plus five group factors. Factor scores were constructed for a random sample of 100 cases and transformed into item scores and their covariances. The following coefficients were obtained:

| | $\rho_{++}$ | $\rho^*$ | $\alpha_s$ |
|---|---|---|---|
| Population value | .819 | .819 | .819 |
| Sample estimate | .906 | .835 | .813 |

A proper evaluation of bias would of course require a great number of samples and a variety of parameter sets, but this example suffices to show that the bias in $\hat{\rho}_{++}$ needs to be considered seriously if item covariances are to be factored. Crossvalidation may be essential. (Some BW examples were based on samples on the order of 100, but they analyzed only matrices of low order.)

In a sufficiently large sample, $\hat{\rho}^* \leq \hat{\rho}_{++} \leq \hat{\rho}_{tt}$. A factor is "common" with reference to a particular set of variables. Enlarging the set of variables entering the factor analysis, as we do in going from $\hat{\rho}^*$ to $\hat{\rho}_{++}$, is likely to change some variance in the initial set from "specific" to "common", and thus to raise communalities. Because $\hat{\rho}^*$ capitalizes on chance (ordinarily to a lesser degree than $\hat{\rho}_{++}$), we cannot be sure that a small-sample $\hat{\rho}^*$ is less than $\rho_{tt}$.

Further inequalities are worthy of note. Just as in the population, the sample $\hat{\alpha}_s \leq \hat{\rho}_{++}$ for the total test or for a part score. (If the part is not stratified, for the part score $\alpha_s$ becomes $\alpha$ and $\rho_{++} = \rho^*$.) Communalities entering $\hat{\rho}^*$ include variance from factors beyond the first centroid of items within the part, whereas only that centroid contributes to the numerator of $\hat{\alpha}$ for the part. Therefore, for the coefficients obtained by considering all parts, $\hat{\alpha}_s \leq \hat{\rho}^*$. Finally, it is likely that $\hat{\rho}_+ < \hat{\rho}^*$ from the same parts. This would follow in large samples from the earlier intuitive argument that $\rho_+$ from sets tends to be less than $\alpha_s$. The inequality might be reversed, however, if scores from a modest sample are divided into a great number of item sets (since the covariance matrix for obtaining $\hat{\rho}_+$ is then large).

Because $\hat{\alpha}_s \leq \hat{\rho}^*$, $\hat{\rho}^*$ tends to overestimate $\rho_s$, the correlation for stratified-parallel tests. The still larger $\hat{\rho}_{++}$ is even less suitable as an alternate-forms coefficient.

## Recognizing Nonindependence

Turn now to the possibility that, while sets of items are experimentally independent, items are not. A familiar example is the reading comprehension test where several questions refer to the same passage. On such questions, two errors can arise from a single haphazard misinterpretation. The misinterpretation can be regarded as random error, but both responses reflect this one departure from the true level of ability.

When Bentler and BW calculated an $\hat{\alpha}$ from the matrix of between-set covariances, they in effect entered the mean of these covariances in diagonal cells of the matrix. If items within sets are experimentally independent (sets being fixed), this is an inappropriate formula, as Lord and Novick (1968, p. 91) pointed out in connection with TOEFL. A factor common to true scores of items in one set and not found in any other set is mistakenly treated as error. It should be noted, however, that sets are sometimes interpreted as random samples from a domain, and then it is appropriate to treat person × set interaction as error (Cronbach, Gleser, Nanda, & Rajaratnam, 1972, p. 236).

If the assumption of independence of items within sets is not accepted, calculating $\hat{\rho}_+$ from covariances of sets is justified. In general, $\hat{\rho}_+$ from smallest independent parts is a superior large-sample estimate of $\rho_{tt}$. Also, in this situation, it is appropriate to average covariances of smallest independent *sets* to obtain diagonal terms for an alpha coefficient.

## The Comrey Example

A complete set of correlations and s.d.'s from 750 cases was supplied by Comrey for his conformity items. Most of these cases entered the analyses in his manual, on which

TABLE 2

Coefficients Obtained for Comrey Scores by Averaging Covariances

| Row | Covariances averaged | Order of matrix[a] | *m*[b] | Coefficient for Quad 2 | Coefficient for test |
|---|---|---|---|---|---|
| 1 | Quads | 5 | 4 | | .823[c] |
| 2 | All duos | 10 | 2 | | .875 |
| 3 | All items | 20 | 1 | | .908 |
| 4 | Duos within quad | 2 | 2 | .78 | .918 |
| 5 | Items within quad | 4 | 1 | .81 | .930 |
| 6 | Items within duo | 2 | 1 | .82 | .937 |

[a] Order of set within which averages are taken; average is based on one less than this number of covariances.

[b] Number of items in smallest subset taken as a variable.

[c] Coefficient evaluated by BW; may be compared with the "floor" of .862 obtained with zeros in diagonal cells of the item-covariance matrix.

BW relied; using all 750 cases changes the various coefficients negligibly. It will be recalled that the 20 conformity items are divided into five sets; the four-item set will be called a "quad", and like-keyed items within a quad a "duo". Although I carried out calculations on all quads, the table is simplified by reporting coefficients for the full test and for Quad 2 only. Each coefficient for Quad 2 was at or near the median of such coefficients for all quads. (Since sampling of persons is not a central concern here, the ^ symbol is omitted from the text and from Table 3.)

Table 2 lists coefficients calculated by averaging item covariances. It was argued earlier that the coefficient in Row 1 is inappropriate, and the coefficient in Row 2 has the same reliance on between-set covariances. Row 3 gives the simple $\alpha$ (expected to be unsuitably low for a stratified test). The $\alpha_s$ coefficients in Rows 4 and 5 place items with unlike keying in the same part, while Row 6 gives $\alpha_s$ for Comrey's narrow strata. This coefficient is greatest because the analysis fully recognizes the two-level stratification.

Table 3 gives results from CMTFA. Rows are ordered by size of coefficient, but row numbers are assigned to correspond to Table 2. In both tables, increase in coefficients is associated with smaller *m*. Holding *m* constant, larger order of the matrix analyzed goes with a larger coefficient in Table 3 and the reverse is true in Table 2. The test coefficient from covariances of duos within the quad has no label, though it is roughly like a $\rho^*$. The two test coefficients labelled $\rho^*$ are subtly different. The one from items within quads is .005 greater; it considers more information and is also subject to more bias.

Greatest interest attaches to the largest values: .951 for $\rho_{++}$, .942 for $\rho^*$ from quads, and .937 for $\alpha_s(=\rho^*$ from duos). In this example, the difference among the coefficients is not impressive, especially when we recall that greater bias goes with greater order in Table 3. Among coefficients for the five quads, $\rho_{++}$ typically exceeds $\alpha_s$ by about .015, and never by as much as .03. It would be hazardous to generalize from data for Comrey's test, constructed by a sophisticated stratification plan guided by a factor analysis of items.

## TABLE 3

### Coefficients Obtained for Comrey Scores by Estimating Communalities

| Row[a] | Covariances factored | Order of matrix | *m*[b] | Coefficient for Quad 2 | Coefficient for test |
|---|---|---|---|---|---|
| 1 | All quads | 5 | 4 | | $\rho_+$ = .850[c] |
| 4 | Duos within quad | 2 | 2 | $\rho_+$ = .78 | .918 |
| 2 | All duos | 10 | 2 | | $\rho_+$ = .929 |
| 6 | Items within duo | 2 | 1 | $\rho^*$ = .82 | $\rho^*$ = .937 |
| 5 | Items within quad | 4 | 1 | $\rho_{++}$ = .83 | $\rho^*$ = .942 |
| 3 | All items | 20 | 1 | | $\rho_{++}$ = .951 |

[a] Numbering corresponds to Table 2.

[b] Number of items in smallest subset taken as a variable.

[c] Coefficient evaluated by BW; see note *c* in Table 2.

Still, the similarity among $\rho_{++}$, $\rho^*$, and $\alpha_s$ (the coefficients that are not plainly inappropriate) is much more striking than their differences.

## Conclusions

1. Where items are independent and the sample of persons is large, CMTFA of items gives a greatest defensible internal-consistency coefficient $\rho_{++}$—a superior lower bound to immediate retest reliability.

2. A coefficient $\hat{\rho}^*$ gives an estimate of immediate retest reliability that is subject to less upward bias.

3. Applying CMTFA to covariances between sets of items does not give a suitable coefficient when items within sets are experimentally independent, and person-set interaction is regarded as a source of true variance.

4. For an investigator concerned with reliability across stratified-parallel forms, the appropriate estimator is $\hat{\alpha}_s$.

### References

Bentler, P. M. (1972). A lower bound method for the dimension-free measurement of internal consistency. *Social Science Research, 1*, 343–357.

Bentler, P. M., & Woodward, J. A. (1980). Inequalities among lower bounds to reliability: With applications to test construction and factor analysis. *Psychometrika, 45*, 249–267.

Bentler, P. M., & Woodward, J. A. (1983). The greatest lower bound to reliability. In H. Wainer & S. Messick (Eds.), *Principals of modern psychological measurement* (pp. 237–253). Hillsdale, NJ: Erlbaum.

Comrey, A. L. (1970). *EITS manual for the Comrey Personality Scales.* San Diego: Educational & Industrial Test Service.

Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. *Psychometrika, 12*, 1–16.

Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: Wiley.

Cronbach, L. J., Ikeda, H., & Avner, R. A. (1964). Intraclass correlation as an approximation to the coefficient of generalizability. *Psychological Reports, 15*, 727–736.

Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement, 25*, 291–312.

Della Riccia, G., & Shapiro, A. (1980). *Minimum rank and minimum trace of covariance matrices.* Ben-Gurion University of the Negev, department of Mathematics, Beer-Sheva, Israel.

Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika, 10*, 255–282.

Jackson, P. W., & Agunwamba, C. C. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous types: I. Algebraic lower bounds. *Psychometrika, 42*, 567–578.

Jackson, R. W. B., & Ferguson, G. A. (1941). *Studies on the reliability of tests* (Bulletin No. 12). Toronto: University of Toronto, Bureau of Educational Research.

Lord, F. M. (1955). Estimating test reliability. *Educational and Psychological Measurement, 15*, 324–336.

Lord, F. M. (1956). Sampling error due to choice of split in split-half reliability coefficients. *Journal of Experimental Education, 24*, 245–249.

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley.

Rajaratnam, N., Cronbach, L. J., & Gleser, G. C. (1965). Generalizability of stratified-parallel tests. *Psychometrika, 30*, 39–56.

ten Berge, J. M. F., Snijders, T. A. B., & Zegers, F. E. (1981). Computational aspects of the greatest lower bound to the reliability and constrained minimum trace factor analysis. *Psychometrika, 46*, 201–213.

Tryon, R. C. (1957). Reliability and behavior domain validity: Reformulation and historical critique. *Psychological Bulletin, 54*, 229–249.

Woodhouse, B., & Jackson, P. W. (1977). Lower bounds for the reliability of the total score on a test composed of nonhomogeneous types: II. A search procedure to locate the greatest lower bound. *Psychometrika, 42*, 579–592.

Zegers, F. E., & Knol, D. L. (1980, June). Chance capitalization in the greatest lower bound to the reliability of a test. Paper presented at the European meeting of the Psychometric Society, Groningen, The Netherlands.

*Manuscript received 9/16/86*
*Final version received 2/10/87*