DOCUMENT RESUME

ED 141 378 TM 006 037

AUTHOR Stenner, A. Jackson; Katzenmeyer, W.G.
TITLE Structural Integrity in Measures of Self Concept.
NOTE 19p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Analysis of Covariance; Bilingual Students; Black Students; Caucasian Students; Chinese Americans; Demography; Elementary Education; *Elementary School Students; Factor Analysis; Minority Group Children; Racial Differences; *Self Concept Tests; Spanish Speaking; Statistical Analysis; Test Bias; Test Construction; *Test Reliability; *Test Validity
IDENTIFIERS *Self Observation Scale (Katzenmeyer and Stenner)

ABSTRACT

Structural integrity of a measure is defined in terms of its replicability, constancy, invariance, and stability. Work completed in the development and validation of the Self Observation Scales (SOS) Primary Level (Stenner and Katzenmeyer, 1973) serves to illustrate one method of establishing structural integrity. The name of each scale of the SOS is a construct: an abbreviated thesis about the nature of the underlying variable which determines the configuration of indicants (questions) comprising the scale. SOS constructs (self acceptance, social maturity, school affiliation, and self security) are analyzed for different student groups. Traditional psychometrics assumes that the same questions measure the same underlying variables in all respondents; the validity of interperson and intergroup comparisons rests on this assumption. However, ample evidence suggests that this assumption is frequently in error. Procedures outlined in this paper provide a method for assessing the validity of this assumption, and a basis for developing reliable and valid instruments. (Author/MV)

STRUCTURAL INTEGRITY IN MEASURES OF SELF CONCEPT

A. Jackson Stenner
National Testing Service, Inc.

W. G. Katzenmeyer
Duke University

Cattell (1964) has suggested the term "structural psychometrics" to refer to a measurement theory grounded in the search for regularity and pattern among indicants. Kaplan (1964, p.55) has defined science as a "search for constancies, for invariants". Structural psychometrics is a search for measures of constructs that evidence relationships among indicants which are "constant and invariant", i.e., the indicants for a construct form a constant, invariant and replicable pattern of relationships. A set of indicants evidencing constancy, invariance and replicability, may be said to possess structural integrity (Stenner & Katzenmeyer, 1975). The term 'structural' is used in its classical sense as "of having to do with, or used in building" and the term 'integrity' refers to "undivided or unbroken condition; wholeness, completeness, entirety".

It is useful to consider a construct as the product of a minitheory or hypothesis about the nature of some organizing influence which has determined the relationships holding among a set of indicants. In psychometrics a set of indicants (e.g., questions) generally identified, either factor analytically or intuitively, are thought to be related to some underlying

variable. Structural integrity implies that the relationships among the indicants (structure) of a particular construct are replicable, invariant, and constant. When the indicants of a construct have a high level of structural integrity, the theory is supported and when the relationships among the indicants for a construct continue to obtain across multiple instances or opportunities, increased support is provided to the inferences (1) that the indicants are measuring some univocal underlying variable, (2) that the pattern of covariance has what Cattell has referred to as "law like status" and (3) that the generation of the proper construct (hypothesis about the cause of the underlying relationship) is of particular importance to theory development. The fact that the pattern of relationships is not perfectly invariant does not negate the value of the construct; given the falibility of the indicants, essential similarity, not identity, is the objective. Kaplan (1964) states: "Laws, in whatever science, group individuals together only on the basis of a similarity in some respect or other, not on the basis of identity (or the group would have only one member). Laws deal with what is common to many, but uniqueness does not imply that nothing is shared with other individuals, only that not everything is common to them. A law requires repeatability, for it formulates a constancy of recurrence; but what recurs is not one and the same instance - for this kind of recurrence is contradictory - but another instance sufficiently like the

former ones to serve the purposes of the generalization of which it is an instance". Thus we invoke the Leibnizian principle of "The identity of indiscernibles" and conclude that whenever two or more indicant patterns are indiscernible with respect to structure, i.e., cannot be differentiated from one another, then the same organizing influence(s) (or mix of influences) is taken to be operating. Two indicant patterns will never be identical but they may be sufficiently similar to permit the generalization that they are the product of the same influence(s).

Construct definition, where definition is viewed as the process of "specification of meaning" (Kaplan, 1964), is taken to be a preeminent objective of educational and psychological research. Construct definition can be usefully considered as embracing two more specific processes, namely intrinsic and extrinsic definition. The former is concerned with identifying the organizing influence through an understanding of the relationships holding among the indicants for the construct and demonstration of the constancy, invariance, and replicability of these relationships. Extrinsic definition, on the other hand, focuses on the relationships between the construct inferred from the pattern, and other constructs. By far the bulk of attention in educational and psychological measurement in the past twenty years has been given to extrinsic definition and, consequently, only sporadically have the structural properties of measures been examined. Ernhart and Loevinger (1969) comment

on the fact that patterns of relations among indicants (although assumed to be invariant in classical measurement theory) may not remain invariant across changes in the demographic characteristics of the sample. They state:

> Methodologically our main finding is that the patterns of relations between items and between tests can be altered by arbitrary choice of demographic variables held constant or allowed to vary .... Psychologists have gradually come to accept that test construction, to be on a professional plane, requires far more than inventing an ingenious set of items. A variety of empirical studies is also mandatory. Results of the present study push that recommendation further. Factorial studies, whether by factor analysis, homogeneous keying, or other method, cannot be performed at a professional level with a small homogeneous sample. Robustness under changes in the demographic character of the sample is a major clue to the psychological reality of and importance of a trait (pp.79-80).

Consistent with the previous discussion, it might be argued that robustness (i.e., stability in the pattern of covariances among the indicants) under changes in the demographic character of the sample is not simply a "clue to the psychological reality of a trait" but rather a statement regarding the validity of a measure. For example, if we accept that similarity in indicant structure across groups of black subjects and white subjects permits the inference that the same organizing influence is operative in the two groups, then a lack of structural invariance may call for rejection of the hypothesis that

the same organizing influence is operative. If the indicant covariance patterns associated with a particular construct are dissimilar between two groups, we infer that the same organizing influence is not being measured in both groups. Thus it follows that the interperative significance of score level differences between the two groups is indeterminate (since we are not measuring the same organizing influence in the two groups). A seldom stated and even less often tested assumption in comparative research which involves psychometric tools is that the same thing is being measured in the groups being compared. To the extent to which indicant structures are found to be dissimilar, this assumption is probably false.

Establishment of structural integrity involves the identification of the characteristic pattern of covariance among the indicants of the construct. A covariance structure evidencing high structural integrity has claim to being a good representation, or measure, of some organizing influence(s). Change in the role of a construct, in theory, may alter the interpretation or extrinsic definition of the construct, but such change has no impact whatsoever on the intrinsic definition of the construct. Thus, a finding bearing on the intrinsic definition of a construct has a direct bearing on the extrinsic definition of the construct while the converse is not the case. For example, a finding that the indicants for a self concept factor "self acceptance" are not structurally invariant across middle SES, Anglo subcultures

and lower SES, Latino subcultures, would restrict the generalizability of the construct and have immediate consequences for both the intrinsic and extrinsic definitions of the construct. Whereas, if the same set of indicants had demonstrated structural integrity, including invariance across American subcultures, and it was found that "self acceptance" predicted science achievement quite well in a sample of sixth grade Anglo children and not at all in a similar sample of Latino children, we would have no cause for questioning our intrinsic definition of the construct. The latter finding might generate the hypothesis that in low SES, Latino subcultures science achievement is not important enough in the reinforcement hierarchy to enter into a child's assessment of himself, whereas, in the middle SES, Anglo subcultures all types of scholastic achievement are prized. Note that this hypothesis impacts the extrinsic definition of the construct, not the intrinsic definition. The fact that the construct "self acceptance" has a different pattern of extrinsic relationships across Anglo and Latino subcultures would not alter the fact that if the indicant structures (i.e., intrinsic relationships) are essentially similar across the two groups, then the same organizing influence is being measured by the indicants within the two groups. Extrinsic definition presupposes intrinsic definition while the reverse does not obtain. That this is the case should be apparent when we realize that an organizing influence may be discovered and measured with indicants evidencing high structural integrity, and yet the measure of the

construct be unrelated to all behaviors considered relevant by contemporary theory.

Four characteristics of a measure are organized under the rubric "structural integrity"; these are: (1) replicability - the extent to which a structure appears in essentially the same form in randomly selected samples of individuals; (2) constancy - the degree to which a structure appears in essentially the same form across the range of observed scores, (3) invariance - the similarity of the structure across selected groups with varying characteristics; and (4) stability - the extent to which the structure is similar across two or more administrations of the measure to the same individuals.

Work completed in the development and validation of the Self Observation Scales (SOS) Primary Level (Stenner and Katzenmeyer, 1973) will serve to illustrate one method of establishing structural integrity. The name of each scale of the SOS (e.g., Social Maturity) is a construct: an abbreviated thesis about the nature of the underlying variable which determined the configuration of indicants (questions) comprising the scale. Cronbach (1971) noted that constructs are "deliberate creations chosen to organize experience into general law-like statements" (p.462). The term 'construct' is particularly appropriate because it connotes human invention. The regularities in human behavior, or specifically, the regularities in children's responses to SOS items, were discovered. The constructs

created to understand these regularities were invented. We invented the label "Social Maturity" and the verbal description of what the term implies to explain a discovered regularity in item configuration.

Table 1 presents the coefficients of replicability, invariance and constancy for each of the scales of the SOS, Primary Level. Table 2 presents a description of the sampling procedure used in each of the studies referenced in Table 1. Since similar factor analytic methodology was used in determining the coefficients of replicability, invariance and constancy, this procedure will be reported in detail only for replicability coefficients. We shall deal with replicability, invariance and constancy in order.

Replicability - Study 1

The data used in this study were obtained from the national norming and validation of the Primary Level of the Self Observation Scales (SOS) (Stenner and Katzenmeyer, 1973). The subjects of the study included first, second and third graders who responded to the 50 items of the (SOS) Primary Form A during April and May, 1973.

A sample of 6,300 cases was divided into four random subsamples (replicates) of 1,575 cases each. The steps involved in the separate factor analyses of each random subsample are described in 1 through 5. Steps 6-9 describe the procedures involved in computing the coefficients of replicability.

1. A matrix of phi coefficients was computed. When a missing datum was encountered, the mean value for that variable was inserted. The percentage of missing data was less than 4%.
2. Squared multiple corrleations were entered as initial communality estimates. Iteration for communalities proceeded until the maximum absolute deviation between iterations dropped below .001.
3. A rotation to the varimax criterion was performed.
4. The orthogonal varimax solution was rotated to maximum oblique simple structure, using the Maxplane criterion (hyperplane width .15).
5. The matrix of loadings of the variables on the factors V(fe) was computed using $V_{fe} = R_v^{-1} V_{fs}$, where R_v^{-1} is the inverse of the matrix of correlations among the variables and V_{fe} is the oblique factor structure (matrix of correlations of factors and variables).
6. Scores on each variable (question) for the total group (6,300 cases) were converted to z scores and factor score estimates (least squares regression estimates) were computed for each subject, on each factor, using the four V_{fe}'s. Since four factors were identified in each of the rotations, this procedure resulted in 16 factor score estimates on each of four factors.

7. Correlations between the estimated and true factor scores were computed (multiple correlation of the estimated scores with the 45 variables of the data matrix, which is also the standard deviation of the estimated factor scores).
8. Correlation coefficients between factor score estimates from each replicate pair (six pairs) were computed. This procedure produced six estimates of the coefficient of replicability for each factor.
9. Coefficients of replicability and confidence intervals associated with these coefficients were obtained in the following manner: Fisher's r to z transformation was performed with each of the six coefficients of invariance obtained for each factor. The means and standard deviations of Fisher z values were obtained and confidence intervals computed ($p < .05$, $p < .01$). The r equivalents of the mean Fisher z value and of the 95 and 99% conficence limit z values were computed.

Study 2 utilized the same procedures as Study 1 but replicates of 700 cases each were utilized in lieu of the larger replicates of Study 1. It may be noted that replicability coefficients are lower in samples of smaller size.

Invariance

In studies 3-8 primary age subjects were divided into groups as indicated in Table 2. Steps 1 through 7 were carried out as indicated above. The correlation coefficients between factor score estimates from a Vfe derived from all subjects were computed for each scale. These are the coefficients of invariance reported in Table 1.

Constancy

The constancy study utilized the same factor analytic methodology as did the invariance studies except that subjects were divided into quartiles with respect to their scores on each scale before the constancy of that scale was checked. This division of subjects into quartiles resulted in an attenuation of the range of the variables of greatest concern and consequently the appropriate correction for this attenuation (Guilford, 1965) was applied to the obtained correlations before the factor analyses were undertaken. From the point of this correction the analysis proceeded as in the computation of coefficients of invariance. A complete sequence of these steps was completed separately for each scale. That is, subjects were first grouped with respect to their quartile on the Self Acceptance scale, then with respect to their score on the Social Maturity scale, etc. This procedure resulted in the constancy coefficients reported in Table 1.

Examination of Table 1 indicates that Self Acceptance is satisfactorily replicable and invariant across sex and race. In both study 3 and 4, black males evidence the lowest invariance on the Self Acceptance scale. The possibility that the lower invariance for black males derives from a cultural influence leading to different interpretation of the items is probably unlikely since structural invariance among black females is not similarly low, and since Mandarin Chinese children responding to a Chinese version, whom we might expect to have major cultural differences, produce higher structural invariance coefficients. The low invariance of Self Acceptance across the Spanish (Study 7) version may be partly due to translation difficulties and/or partly due to a confounding of Self Acceptance and School Affiliation in the Latino culture.

One explanation for the lower constancy coefficient in the first quartile on the Social Maturity and School Affiliation scales is that a ceiling effect exists which restricts the variances of the items which load these scales sufficiently to distort the structure. Subsequent research with additional items at the Intermediate Level, Grades 4-6, supports this conclusion.

Summary

A low coefficient of replicability on any scale suggests that such patterns of covariance as have been discovered in individual random replicates differ from each other. There is no clearly defined pattern of covariance among these indicants which is

common to random subsamples from this population. Replicability is prerequisite to useful structural integrity. If replicability cannot be demonstrated, no further analyses should be undertaken. It should be observed that the number of cases required to achieve satisfactory invariance may vary with the quality of the indicants, and with the domain being sampled. In the self concept domain the authors found the ratio of cases to variables needed to be about 7 to 1 before satisfactory replicability coefficients could be obtained.

A low coefficient of invariance suggests that the same underlying variable may not be being measured in the two groups across which the coefficient of invariance is low. This means that comparisons of scores earned on that particular scale by members of the different groups are not comparable. Items which contribute to low invariance coefficients may be identified and eliminated, or they may be used in an attempt to understand the nature of the differences between the groups in question.

Low coefficients of constancy suggest that the underlying variable which is being measured may not be the same for respondents in the quartile(s) having the low constancy coefficient.

Low replicability found with random subsamples of adequate size is incurable with the item set available. No clearly defined patterns of covariance exist among the indicants. A new or augmented set of indicants is needed. Low invariance coefficients, where sample sizes are adequate, may reflect characteristically different interpretations of the indicants between

responding groups, essential differences in the structure of the construct between the groups, or an interaction between group membership and score level where constancy coefficients are also low on the scale in question. Low constancy coefficients generally are indicative of one or more psychometric problems. Low constancy in the bottom quartile suggests a floor effect; low constancy in the top quartile, a ceiling effect. Response bias may account for low constancy in any quartile. Once the source of the low constancy coefficient has been identified, steps can be taken to eliminate the lack of constancy.

Traditional psychometrics assumes that the same questions measure the same underlying variables in all respondents. The validity of interperson and intergroup comparisons rests on this assumption. Ample evidence exists to suggest that this assumption is frequently in error. The structural integrity procedures outlined in this paper provide a method for assessing the validity of this assumption, and provide a basis for the development of instruments which possess structural integrity.

Table 1

Structural Integrity Matrix
Primary SOS (Form A)

	SAC	SM	SAF	SS
Replicability				
Study #1 N = 1575 per replicate	.96	.98	.99	.97
Study #2 N = 700 per replicate	.91	.98	.98	.94
Invariance				
Study #3				
White males	.97	.99	.99	.99
Black males	.79	.96	.99	.99
White females	.98	.98	.96	.98
Black Females	.87	.97	.99	.97
Study #4				
White males	.91	.96	.99	.97
Black males	.84	.81	.99	.91
White females	.94	.96	.99	.97
Black females	.89	.96	.99	.97
Study #5 Spanish bilinguals responding to English	.88	.83	.93	.96

SAC: Self Acceptance
SM: Social Maturity
SAF: School Affiliation
SS: Self Security

Table 1 (continued)
Structural Integrity Matrix
Primary SOS (Form A)

	SAC	SM	SAF	SS
Invariance				
Study #6 Chinese children responding to Mandarin version	.80	.50	.75	.95
Study #7 Latinos responding to Spanish version	.45	.78	.93	.93
Study #8				
First graders	.97	.92	.99	.99
Second graders	.99	.96	.98	.99
Third graders	.98	.96	.99	.99
Constancy				
N=1000/Quartile				
First quartile (highest	.83	.52	.43	.91
Second quartile	.96	.90	.93	.94
Third quartile	.82	.92	.95	.94
Fourth quartile	.95	.87	.99	.97

Table 2

Comments on Studies from Which
Structural Integrity Coefficients Reported
in Table 1 Were Obtained

1. A national sample of 6,300 cases was divided into four random subsamples (replicates) of 1,575 cases each. The factor analysis procedure is described in Katzenmeyer & Stenner (1975).

2. A national sample of 2,800 cases was divided into four random subsamples of 700 cases each. The same replicability procedure used in (1) above was applied.

3. A national sample of first, second, and third graders was divided into four sex-race groups as follows: 1,265 white males; 732 black males; 1,201 white females; and 702 black females. Each group was factor analyzed according to the procedure described in Stenner & Katzenmeyer (advance publication #5).

4. A national sample of 2,800 cases was divided into four race-sex groups of 700 cases each. The uniformity procedure described in (3) above was applied.

5. A sample of 4,295 first, second, and third grade bilingual Latino children from a large midwestern metropolitan school district was administered the Primary SOS in English. The uniformity coefficients represent the match between the national norming sample factor structure and the bilingual Latino sample factor structure.

6. A sample of 183 Chinese children, ages 6-8, participating in a bilingual education program in a midwestern metropolitan school system was administered a Mandarin Chinese version of the SOS. The responses were factor analyzed and the resulting factor structure was compared with the national norming sample factor structure.

7. A national sample of 6,300 cases (same sample as (1) above) was divided into quartiles on the basis of scores on each factor. Thus, scores on Self Acceptance were ordered from highest to lowest, then the sample was divided into quartiles. The quartile groups were then factor analyzed and the resulting factor structures were compared with the national norming sample factor structure. The above procedure was repeated for each of the other three Primary Self Observation Scales.

8. A sample of 2,151 Latino children, ages 5-8, participating in a bilingual education program in a midwestern metropolitan school system was administered a Spanish version of the SOS. The responses were factor analyzed and the resulting factor structure was compared with the national norming sample factor structure.

Bibliography

Cattell, R. B. "Validity and Reliability: A Proposed More Basic Set of Concepts." Journal of Educational Psychology, 1964, 55, 1, p.1-22.

Cronbach, Lee J. "Test Validation," in R.L. Thorndike Educational Measurement, Washington: American Council on Education, 1971, p.443-507.

Ernhart, Claire B. and Jane Loevinger. "Authoritarian Family Ideology: A Measure, Its Correlates, and Its Robustness." Multivariate Behavioral Research Monographs, MBR Monograph No. 69-1, 1969.

Guilford, J. P. Fundamental Statistics in Psychology and Education. New York: McGraw-Hill, 1956.

Kaplan, Abraham. Conduct of Inquiry. San Francisco: Chandler Publishers, 1964.

Katzenmeyer, W. G. and A. Jackson Stenner. "Strategic Use of Random Subsample Replication and a Coefficient of Factor Replicability." Educational and Psychological Measurement, 1975, 35, 19-29.

Katzenmeyer, W. G. and A. Jackson Stenner. Self Observation Scales - Primary Level, Durham, North Carolina: National Testing Service, Inc., 1973.