DOCUMENT RESUME

ED 223 704 TM 820 851

AUTHOR Reuman, David A.; And Others
TITLE Measurement Models for Thematic Apperceptive Measures of the Achievement Motive.
SPONS AGENCY National Inst. of Mental Health (DHHS), Rockville, Md.
PUB DATE 27 Aug 82
GRANT MH-14618-05
NOTE 32p.; Paper presented at the Annual Meeting of the American Psychological Association (90th, Washington, DC., August 27, 1982).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Academic Achievement; *Achievement Need; Adults; Error of Measurement; Job Satisfaction; *Models; Test Theory; *Test Validity; *True Scores
IDENTIFIERS *Classical Test Theory; Nomological Network; *Thematic Appreciation Test; Theoretical Analysis

ABSTRACT

According to classical test theory, the presence of random measurement error in a psychological test has important implications for validation studies. The more comprehensive application of classical test theory in construct validation is distinguished from that in criterion-oriented validation. Critics of thematic apperceptive measurement of the achievement motive have often blurred this distinction and have consequently failed to appreciate the construct validity of this motive measure. Using explicit true score measurement models of theoretical constructs, evidence for the construct validity of the achievement motive in a representative sample of adult males in America (n=508) is described. Evidence is presented that the achievement motive construct is related in theoretically expected ways to constructs of academic achievement and work satisfaction, despite the presence of substantial random measurement error in thematic apperceptive measures of the motive. Evidence for the discriminant validity of story content from story length in the thematic apperception test is also presented in this nomological network. (Author/PN)

Measurement Models for Thematic Apperceptive Measures of the Achievement Motive

David A. Reuman, Duane F. Alwin, and Joseph Veroff

The University of Michigan

Abstract

According to classical test theory, the presence of random measurement error in a psychological test has important implications for validation studies. We distinguish the more comprehensive application of classical test theory in construct validation from that in criterion-oriented validation. Critics of thematic apperceptive measurement of the achievement motive have often blurred this distinction and have consequently failed to appreciate the construct validity of this motive measure. Using explicit true score measurement models of theoretical constructs, we present evidence for the construct validity of the achievement motive in a representative sample of adult males in America ($N = 508$). We present evidence that the achievement motive construct is related in theoretically expected ways to constructs of academic achievement and work satisfaction, despite the presence of substantial random measurement error in thematic apperceptive measures of the motive. Evidence for the discriminant validity of story content from story length in the thematic apperception test is also presented in this nomological network.

This research was made possible by a grant from the National Institute of Mental Health, MH 14618-05. We appreciate the comments of James A. McRae, Jr. and Warren T. Norman on earlier versions of this paper.
Address correspondence to: David A. Reuman, Department of Psychology, The University of Michigan, Ann Arbor, MI 48109.

This paper was presented at the 90th Annual Convention of the American Psychological Association, Washington, DC; August 27, 1982.

The thematic apperception test (TAT) and the concept of need for achievement (n Achievement) have a shared history which dates back more than 40 years (Murray, 1938). As a theoretical construct, the achievement motive represents a broad, affectively-toned disposition to compete with a standard of excellence. The motive construct is expected to influence behavior in an extensive set of situations where individuals experience pride in accomplishment and disappointment in failure (Atkinson, 1958; Atkinson and Feather, 1966; Atkinson and Raynor, 1974; McClelland, Atkinson, Clark, and Lowell, 1953; McClelland and Winter, 1969; Veroff and Feld, 1970).

As a technique for measuring the achievement motive, the TAT has generally shown modest test-retest and internal consistency reliability (Murstein, 1963). Critics have repeatedly contended that the low reliability of thematic apperceptive measures precludes the demonstration of validity for the achievement motive (Entwisle, 1972; Fineman, 1977; Klinger, 1966). This argument is based on a fundamental principle of classical test theory (Lord and Novick, 1968); namely, that the validity of a test with respect to any criterion cannot exceed the index of reliability of the test (i.e., the square root of reliability). In other words, high reliability is a necessary but not a sufficient condition for high criterion validity. Criterion validity includes both predictive validity and concurrent validity, but differs in important respects from construct validity (Cronbach and Meehl, 1955). Successful criterion validation depends only on the degree of observed covariation between predictor variables and a criterion variable.

Some writers have argued that the achievement motive as assessed by the TAT lacks convergent validity (e.g., Fineman, 1977) and interpret this failure in light of the unreliability of the TAT. Similarly, writers have argued that the achievement motive as assessed by the TAT lacks discriminant validity. Entwisle (1972), for example, has argued that thematic apperceptive measures of the achievement motive are confounded with TAT story length, "conceived as an indicator of verbal achievement and probably also of academic socialization" (p.387). This claim is based on the observation that TAT n Achievement and TAT story length covary positively, and observed zero-order correlations between TAT story length and grades in school are nominally larger than zero-order correlations between TAT n Achievement and grades in school. Entwisle suggests that evidence of the discriminant validity of TAT n Achievement with respect to TAT story length is not forthcoming, because measurement error in thematic apperceptive measures of the achievement motive severely limits the capacity of the test to correlate with any criterion.

These critics have evaluated the convergent and discriminant validity of TAT n Achievement by examination of criterion validity coefficients relating observed test scores. Such criterion validity coefficients confound measurement error in the observed variables with the true relationships between theoretical constructs. We believe that these issues of convergent and discriminant validity are inadequately addressed by examination of criterion validity coefficients alone and ought to be placed in the more comprehensive context of construct validation (Alwin, 1974). The search for convergence among tests presupposes that no single criterion is fully adequate to define the underlying dimension of interest; this presupposition is precisely what Cronbach and Meehl (1955, p.282) define as construct validation. Similarly, discriminant validity presupposes the logic of construct validation. In order to justify why two observed tests ought to show discriminant validity, we require laws in a nomological network which relate (a) the observed tests to other observed tests, (b) theoretical constructs to the observed tests, and (c) these different theoretical constructs to each other (Cronbach and Meehl, 1955, p.290). Clearly, construct validation goes beyond criterion validation in requiring a theoretical context for evaluating validity.

We believe that critics of TAT n Achievement have selectively invoked the logic of classical test theory by restricting their attention to criterion validity coefficients. We apply classical test theory in the fuller context of construct validation by distinguishing between "observed" (criterion) and "true" (construct) validity coefficients, the latter having been disattenuated for measurement error. We propose to examine evidence for the construct validity of TAT n Achievement in an investigation of the discriminant validity of the achievement motive and verbal achievement. We predicate our examination of relationships involving these theoretical constructs on justifiable measurement models of the constructs.

The achievement motive is conceived as an affective disposition to find those situations attractive in which personal performance is evaluated against some standard of excellence. Because performance evaluation is typically made salient in academic situations, researchers (e.g., Entwisle, 1972) assume that the achievement motive ought to covary positively with academic achievement, measured by grades or years of schooling completed, for instance. Because schools invariably demand verbal achievement, it is reasonable to suppose that greater verbal achievement, measured by TAT story length, will also be associated with higher academic achievement. On the other hand, we expect that affective concern over doing well will be systematically related to behavior in many situations in which verbal skills are only incidentally required. We hypothesize that work settings fall into this class of situations. Affective reactions to work ought to be related to the achievement motive, but not to verbal achievement. If the achievement motive and verbal achievement (both measured by the TAT) are found to be related to a relevant construct (e.g., work satisfaction) in different ways, they should be conceptualized as distinct constructs, even though they may covary positively with each other. This would represent evidence for the construct validity of TAT n Achievement (Cronbach and Meehl, 1955).

Selecting a measurement model for TAT n Achievement

Classical true score theory partitions the variation in a measured variable (X) into two orthogonal components--true score variance and random measurement error variance (Lord and Novick, 1968). According to the classical true score model, a true score is a hypothetical expected value of X over repeated measurements, and true score variance is reflected in the covariance among different attempts to measure the same underlying construct. The random error component of this model is defined as the difference between observed and true scores, and according to the model, random error variance is reflected in the difference between the variances of true and observed scores. Random measurement error is uncorrelated with true score variation and with variation in other error scores.

The classical true score model of measurement error has provided psychologists with methods for estimating the reliability of composite measures, that is, for estimating the proportion of observed variance that is considered non-error variance. Traditionally, coefficient alpha (Cronbach, 1951) has been heavily relied upon as an estimate of internal consistency reliability. Certainly, this has been the case in the evaluation of TAT n Achievement (e.g., Entwisle, 1972). Despite its popularity, alpha may not always be the most appropriate coefficient for reliability estimation. It can be shown that unless measures demonstrate tau-equivalence (or essential-tau-equivalence) (see Lord and Novick, 1968, p.90), coefficient alpha will under-estimate the reliability of composite measures. Other, more general, reliability formulations are appropriate for estimating internal consistency reliability of composites when the assumption of tau-equivalence cannot be met (Joreskog, 1978). Recently developed methods of confirmatory factor analysis make it possible to evaluate the appropriateness of various measurement models for a particular set of measures (Joreskog, 1974). In this paper we demonstrate empirically the conclusion that others have derived analytically, that when a tau-equivalent measurement model does not fit the data well, coefficient alpha under-estimates the reliability of the composite measure.

It therefore becomes important in reliability estimation to consider the appropriateness of the measurement model underlying the coefficient used. The most general linear model for estimating the reliability of composite measures is Joreskog's (1978) congeneric measures model. All other models for reliability estimation represent restrictions on this model. This model does not assume tau-equivalence, nor does it place any constraints on the variances of errors, as in the case of parallel measures (see Alwin and Jackson, 1979; Joreskog, 1978). In the following discussion we demonstrate how such differing models for measurement error may be specified within Joreskog's "analysis of covariance structures" framework and how the fit of such models to empirical data can be examined.

Consequences of modeling measurement error for validity estimation

According to classical test theory, the correlation between two theoretical variables will be greater than the correlation between two measured indicators of those theoretical variables whenever random measurement error is present in the indicators. If the magnitude of random measurement error is known or can be estimated, the classical test theorist may disattenuate the correlation between the measured indicators. The disattenuated correlation is treated as a standard for inferring the "true" relationship between the theoretical constructs (Lord and Novick, 1968, p.69). In fact, the objective of determining such true relationships between theoretical constructs originally prompted the development of test theory in psychology.

The pattern of relationships among measured indicators may look quite different from the pattern of relationships among those unmeasured constructs which determine variation in the measured indicators. Suppose constructs A and B are related to a third construct, C, to exactly the same extent. Given the principles of classical test theory, if the indicators of A are substantially more reliable than the indicators of B, we would observe greater covariation between the indicators of A and C than between the indicators of B and C. Analysts who consider only observed covariation might fail to conclude that the constructs, A and B, are equally related to C. Alternatively, suppose that the true covariation between constructs X and Z is significantly greater than the true covariation between constructs Y and Z, but the indicators of X are much less reliable than the indicators of Y. Given the principles of classical test theory, the observed relationship between indicators of X and Z would be attenuated to a greater extent than the observed relationship between indicators of Y and Z. The observed relationship between X and Z indicators might be not significantly different from the observed relationship between Y and Z indicators. In this case, analysts who consider only observed covariation might fail to conclude that the constructs, X and Y, are differentially related to construct Z.

Failure to move beyond simple consideration of covariation among observed variables has led critics of thematic apperceptive methods to misinterpret evidence relevant to the validity of the achievement motive as a theoretical construct. Entwisle (1972), for instance, acknowledged that TAT n Achievement is much less reliable than TAT story length, without actually going the next step of incorporating this knowledge into the discriminant validation endeavor. Instead, the comparison was based on the magnitude of criterion validity coefficients, which confound measurement error and true covariation between constructs. We believe TAT n Achievement has been particularly vulnerable to such interpretive errors, precisely because its reliability tends to be much lower than that of indicators of other constructs. Here we illustrate an approach that overcomes these errors.

Method

Sample

The data reported here originate in a 1976 national survey designed to obtain multiple measures of well-being and role reactions (Veroff, Douvan, and Kulka, 1981). The respondents, drawn from randomly sampled households within the contiguous United States, were chosen to represent Americans over the age of 21. A random subsample (N = 1,498) was administered the TAT near the outset of each interview (see Veroff, Depner, Kulka, and Douvan, 1980, for further description of the survey methods used).

Measures

Respondents told imaginative stories to six pictures, selected to depict a variety of life situations with which respondents would presumably be familiar. Standard TAT prompts used in laboratory studies of motivation helped respondents develop a story plot for each picture. Stories were transcribed verbatim by the interviewers and later coded for achievement imagery and its subcategories using the content analysis procedures developed by McClelland, Atkinson, Clark, and Lowell (1958). Veroff et al. (1980) report satisfactory levels of interscorer coding reliability. A total protocol for a respondent was judged inadequate for assessing strength of the achievement motive if any story contained non-imaginative responses to more than one prompt or if more than three stories were missing one prompt each. On this basis adequate protocols are available for 508 males and 700 females in the survey. Here we analyze data for the 508 males with adequate TAT protocols.

For factor analytic reasons, only TAT items which were not seriously skewed (skewness < 2) were considered here. The four pictures yielding such response distributions for males are (given with their actual order of presentation in parentheses):

(1) Two men (inventors) in a shop working at a machine.

(2) Four men seated at a table with coffee cups. One man is writing on a sheet of paper.

(4) Man at drafting table.

(5) Conference group - seven men variously grouped around a conference table.

The education construct considered here is the amount of formal schooling each respondent received. The indicator of this construct is coded (1) if the respondent completed only grade school, (2) if the respondent completed some high school, (3) if the respondent completed high school, (4) if the respondent completed some college, and (5) if the respondent completed college.

Five indicators of intrinsic satisfaction at work were selected from the 1976 survey. These include:

(1) How much satisfaction (have you gotten/would you get/did you get) out of work at a job? [Coded (1) No satisfaction, (2) Little satisfaction, (3) Some satisfaction, (4) Great satisfaction.];

(2) Here are some things that might describe a person's job. Please tell me how true each is of your (main) job, using one of the answers on this card. The first one is: The work is interesting. [Coded (1) Not at all true, (2)

Not very true, (3) Somewhat true, (4) Very true.];

(3) [This indicator was preceded by the same preface used for indicator 2.] I am given a chance to do the things I do best. [Coded the same as indicator 2.];

(4) How much (has/would/did) work at a job (led/lead) to (the most important value) in your life? [The interviewer substituted that value which the respondent had just selected as most important from a list of nine life values proposed by Rokeach (1973).] [Coded (1) Very little, (2) A little, (3) Some, (4) A lot, (5) A great deal.];

(5) Taking into consideration all the things about your job, how satisfied or dissatisfied are you with it? [Coded (1) Neutral, ambivalent, or dissatisfied, (2) Satisfied, (3) Very satisfied.].

Except for the last item, these indicators used a forced choice response format. Since three of these measures were asked only of currently employed respondents, the number of cases for analyses involving work satisfaction indicators is 413.

Specification of measurement models

In order to select a measurement model, we begin with a variance-covariance matrix calculated from the four non-skewed TAT n Achievement items.[1] We attempt to identify an hypothesized covariance matrix, Σ, in which elements are equal to their corresponding elements in the observed variance-covariance matrix, $\underset{\sim}{S}$. Σ is given by the factor analytic model:

$$\Sigma = \Lambda \; \Phi \; \Lambda' + \Psi^2 ,$$

where Λ is a matrix of factor pattern coefficients relating the $\underline{p}$ measured variables to $\underline{k}$ unmeasured common factors; Φ specifies the variance-covariance matrix relating the $\underline{k}$ common factors; and Ψ^2 is a variance-covariance matrix of $\underline{p}$ disturbance terms (Joreskog, 1969, 1974). We have assumed, without loss of generality, that all factors are centered (i.e., have mean zero) in the analyses that follow. In analysing variance-covariance matrices of TAT data with COFAMM (Joreskog and Sorbom, 1976), we have specified four models:

(1) the <u>null model</u> implies $\underline{k} = \underline{p}$; $\Lambda = \underset{\sim}{I}$; Φ is a diagonal matrix with certain elements constrained to be equal[2]; and $\Psi^2 = \underline{0}$;

(2) the <u>parallel measures model</u> implies that indicators must be univocal (i.e., load only on one factor), and that elements in Λ are constrained to be equal for all indicators of a factor; diagonal elements of Φ are set at 1.0, and off-diagonal elements of Φ are free (i.e., estimated); and Ψ^2 is diagonal with elements constrained to be equal within each set of indicators of a factor;

(3) the <u>tau-equivalent measures model</u> implies that indicators must be univocal, and that elements in Λ are constrained to be equal for all

indicators of a factor; diagonal elements in Φ are set at 1.0, and off-diagonal elements in Φ are free; and Ψ^2 is diagonal and free;

(4) the <u>congeneric measures model</u> implies that indicators must be univocal, and that elements in Λ are estimated for all indicators of a factor; diagonal elements in Φ are set at 1.0, and off-diagonal elements in Φ are free; and Ψ^2 is diagonal and free.

In general, in order to identify uniquely any of these models it is necessary to impose $\underline{k}^2$ restrictions on the elements in Λ and/or Φ. In most cases this may be accomplished by (1) fixing one element in each column of Λ to unity and $\underline{k}$ - 1 elements in each column to zero, or (2) fixing the diagonal of Φ to unities and $\underline{k}$ - 1 elements in each column of Λ to zero. These are necessary but not sufficient conditions for identification (Joreskog, 1979).

The goodness of fit of various measurement models we specify may be evaluated statistically using the methods of maximum-likelihood confirmatory factor analysis. The likelihood-ratio statistic is distributed as X^2 with $\underline{q}$ degrees of freedom, and it is possible to test the hypothesis that a particular specification of values for the free parameters in Λ, Φ, and Ψ^2 will reproduce the population variance-covariance matrix, Σ. The degrees of freedom associated with a model is in general equal to $.5\ \underline{p}\ (\underline{p} + 1) - \underline{t}$, where $\underline{t}$ is the number of parameters of the model that are unconstrained. A nonsignificant $\underline{p}$-value associated with a model implies that the hypothesized factor analytic model generates a matrix of estimated variances and covariances which does not differ significantly from the observed variance-covariance matrix. However, one of the problems with the simple application of these statistical inference tools is that the value of X^2 is influenced by sample size. In sufficiently large samples the value of X^2 may cause one to reject non-trivial models on statistical grounds (Joreskog, 1969). Therefore, as suggested by Joreskog (1978), we also compare changes in the ratio of X^2 to degrees of freedom ($X^2/\underline{df}$) as we compare models. As the $X^2/\underline{df}$ ratio of a model approaches 1.0, fit improves. In addition we make use of Bentler and Bonett's (1980) normed fit index to scale the improvements in fit in standardized units.

Results

Selecting a measurement model for TAT n Achievement

Table 1 displays measures of fit for true score measurement models and their appropriate null model, when these models are applied to the observed variance-covariance matrix of TAT n Achievement indicators for 1976 men. It is evident that the null model and the parallel measures model may be rejected since their $X^2/\underline{df}$ ratios are substantially greater than 1.0, and their p-values indicate extremely poor fit. The tau-equivalent measures model provides a highly significant improvement in fit over the parallel measures model (difference X^2 (3) = 36.840, $p < .001$); the congeneric measures model does not improve significantly upon the tau-equivalent measures model (difference X^2 (3) = 3.792, $p > .250$). Finally, we infer from its $X^2/\underline{df}$ ratio and p-value that the tau-equivalent measures model fits the data very well. These observations lead us to prefer the tau-equivalent measures model in accounting for the relationships among the four indicators of the achievement motive for men in the sample.

Table 1 here

Since these true score models may be arranged hierarchically, we might also determine a preferred model from an examination of normed incremental fit indices (Bentler and Bonett, 1980, p.599). The Bentler-Bonett normed fit indices, calculated from Table 1, are $\hat{\Delta}_{12} = .492$, $\hat{\Delta}_{23} = .452$, and $\hat{\Delta}_{34} = .046$. These incremental indices show smaller improvements in fit, as the hierarchical step-up comparison of two models proceeds. The overall normed fit index for our preferred tau-equivalent measures model is $\hat{\Delta}_{13} = .944$, an acceptable value by Bentler and Bonett's standards (1980, p.600).

Figure 1 displays parameter estimates of the preferred measurement model for 1976 male TAT _n_ Achievement. A reliability coefficient for the indicators of the achievement motive in this tau-equivalent measures model (Alwin and Jackson, 1979, p.96) is .356. This internal consistency coefficient is quite similar in magnitude to the median internal consistency reliabilities reported in psychometric reviews of TAT _n_ Achievement (Entwisle, 1972; Fineman, 1977). Unlike previous reliability estimates, however, our estimate is derived from an explicit, empirically justified measurement model.

Figure 1 here

Relationships among TAT n Achievement, TAT story length, and education

The zero-order correlation between the unweighted sum of the four non-skewed TAT _n_ Achievement items and education for 1976 male respondents is .151 (_df_ = 1, 503; _p_ = .0007); the zero-order correlation between the unweighted sum of the story length measures for these same four TAT items and education is .193 ($df = 1, 503$; $p < .0001$). In a multiple regression analysis with education as the dependent variable, the standardized partial regression coefficients for TAT n Achievement and story length are .100 ($p = .024$) and .157 ($p = .0004$), respectively. This pattern of criterion validity coefficients replicates Entwisle's (1972) finding that story length is nominally the more potent predictor of an educational attainment criterion.

Believing the causal relationships among TAT n Achievement, TAT story length, and education to be ambiguous in our cross-sectional survey design, we proceed to determine the correlation coefficients (rather than structural equation parameters) relating these three constructs, now taking measurement models for the constructs into consideration. Using confirmatory factor analysis, we posit several models that attempt to reproduce the observed variance-covariance matrix relating the four indicators of n Achievement, four indicators of story length, and the single indicator of education.

Table 2 displays measures of fit for these models. Model 1 specifies the null model in which all variance associated with each indicator is random error variance, and error variances are constrained to be equal for indicators of the achievement motive and for indicators of story length. As shown in Table 2, this hypothesized null model should be rejected since its X^2/df is substantially greater than 1.0, and its p-value is highly significant. Models 2 through 6 specify true score measurement models for the n Achievement and story length factors which are hypothesized to underlie the observed indicators. Since only one indicator of the education factor is included in the analysis, and therefore measurement models for that factor cannot be estimated from these data, we impose the assumption that 6.7 percent of the variance in the indicator is error variance, based on prior research which has estimated the reliability of such an indicator (Wheaton, Muthen, Alwin, and Summers, 1977, p.112).

Table 2 here

Contrasts among hierarchically nested measurement models lead us to posit Model 4 as the preferred model for these data.[3] Model 4 specifies that the n Achievement indicators are congeneric, and the story length indicators are tau-equivalent. The model fits the data very well, as shown by its $X^2/\underline{df}$ equal to 1.084 and its nonsignificant p-value. Bentler and Bonett's normed fit index for Model 4 in Table 2 is .980. Model 4 is preferred because it is a significant improvement upon more restrictive models (Models 5 and 6) and is not significantly improved upon by a less restrictive model (Model 2).

In this preferred model (see Figure 2), we note that the correlation between n Achievement and education, which has been disattenuated for the unreliability of the indicators of those constructs, is .256 (SE = .073). The disattenuated validity correlation between the story length and education factors is .210 (SE = .046). The relative magnitude of these "true" validity coefficients implies that TAT n Achievement is in fact as potent as story length in predicting educational attainment. The relative magnitude of the "true" validity coefficients is the reverse of the relative magnitude of the "observed" validity coefficients. These results call into question the conclusions drawn by Entwisle (1972).

Figure 2 here

We note finally that the preferred measurement model for the achievement motive indicators is influenced by the sample of indicators of other constructs selected for analysis. Given indicators of the achievement motive, story length, and education, we now prefer a congeneric measures model for the achievement motive indicators. Given only indicators of the achievement motive earlier, we preferred a tau-equivalent measures model. We interpret this change of preferred models to imply that these indicators of the achievement motive are differentially related to the indicators of other constructs, although they are equally related to the achievement motive construct itself. The reliability coefficient derived from the congeneric measures model of the achievement motive indicators in Figure 2 is .383, similar to that obtained earlier in Figure 1.[4] The reliability coefficient derived from the tau-equivalent measures model of TAT story length in Figure 2 is .914. The fact that the indicators of this construct are so much more reliable than indicators of the achievement motive construct implies that observed validity coefficients of the story length indicators will be changed less when disattenuated for measurement error, compared to observed validity coefficients of the achievement motive indicators.

<u>The discriminant validity of TAT n Achievement and story length</u>

Table 3 displays fit indices for several confirmatory factor analytic models which attempt to reproduce the observed variance-covariance matrix of relationships among indicators of the achievement motive, story length, and intrinsic work satisfaction.[5] It is clear from its fit indices that the null model may be rejected. Systematic contrasts between measurement models for the three constructs lead us to prefer Model 4, which hypothesizes that the indicators of the achievement motive and work satisfaction are congeneric[6], and the indicators of story length are tau-equivalent, consistent with the findings presented in Table 2. Model 4 fits reasonably well ($X^2/\underline{df}$ = 1.362; $\underline{p}$ = .0278). Bentler and Bonett's normed fit index for Model 4 in Table 3 is .953. Model 4 is a significant improvement upon more restrictive models (Models 5 and 6) and is not significantly improved upon by a less restrictive model (Model 2).

Table 3 here

The correlations among the three theoretical variables are displayed in Figure 3. The estimated correlation between the achievement motive construct and the work satisfaction construct is .277 ($\underline{SE}$ = .086); the estimated correlation between the story length construct and the work satisfaction construct is -.011 ($\underline{SE}$ = .057). These correlations are significantly different ($\underline{p}$ < .002), demonstrating the discriminant validity of the achievement motive construct and the story length construct with respect to the work satisfaction construct.

Figure 3 here

Conclusions

Through the systematic application of principles of classical test theory, we have provided evidence here for the construct validity of the achievement motive. Contrary to the claims of Entwisle (1972), our analysis suggests that TAT n Achievement and TAT story length are not equivalent constructs. The achievement motive is expected to influence behavior in an extensive set of situations where individuals compete with a standard of excellence. TAT story length reflects a more localized construct, verbal achievement.

We have differentiated measured variables used as indicators of a theoretical construct from the unmeasured theoretical construct itself. The problem of estimating relationships between a theoretical variable and measured indicators of that theoretical variable, based on sample observations of covariation among the indicators, is conceived here as a reliability issue. The problem of estimating the relationship between two (or more) theoretical variables, based on sample observations of covariation between indicators of each construct, is conceived as a validity issue. Critics of thematic apperceptive methods have failed to distinguish between criterion validity coefficients and validity coefficients that show the relationships between theoretical constructs (i.e., validity coefficients which have been disattenuated for measurement error in indicators of the constructs). As a consequence of this confounding, critics of TAT n Achievement have reached particularly mistaken conclusions, precisely because random measurement error is so much more substantial in thematic apperceptive indicators of the achievement motive than in indicators of other constructs.

By predicating our analysis of the validity of the achievement motive on explicit true score measurement models, we reach estimates of relationships among theoretical constructs which are more theoretically justifiable than those which have been reached by critics of thematic apperceptive methods. Although we believe our approach improves the interpretation of evidence bearing on theoretical propositions, we make no claims for improving the practical utility of the TAT, for instance, as an instrument for selecting individuals into motivational intervention programs. Entwisle's (1972) criticisms of TAT n Achievement are in part directed at such practical limitations of unreliable measures. We would argue simply that questions concerning the theoretical status of TAT n Achievement need not be answered in the same way as questions concerning its practical status.

Future research may refute our choice of a model that specifies that a large proportion of the variance in TAT n Achievement items is due to random measurement error. Motivational theorists (Atkinson, Bongort, and Price, 1977; Reuman, in press) have hypothesized that this "random error variance" is in fact systematically related to motivational dynamics. Our present measurement framework does not address such hypotheses directly. We can at best assert that we have no evidence here that true score measurement models should not be applied. As long as many researchers continue to interpret low correlations among indicators of a construct as a sign of the presence of random measurement error, we would at least recommend that validity coefficients be estimated in conjunction with empirically justifiable measurement models for theoretical constructs.

References

Alwin, D. F. Approaches to the interpretation of relationships in the multitrait-multimethod matrix. In H. L. Costner (Ed.), Sociological methodology 1973-74. San Francisco: Jossey-Bass, 1974.

Alwin, D. F., & Jackson, D. J. Measurement models for response errors in surveys: Issues and applications. In K. F. Schuessler (Ed.), Sociological methodology 1980. San Francisco: Jossey-Bass, 1979.

Atkinson, J. W. (Ed.). Motives in fantasy, action, and society. Princeton, N. J.: Van Nostrand, 1958.

Atkinson, J. W., Bongort, K., & Price, L. H. Explorations using computer simulation to comprehend thematic apperceptive measurement of motivation. Motivation and Emotion, 1977, 1(1), 1-27.

Atkinson, J. W., & Feather, N. T. (Eds.). A theory of achievement motivation. New York: Wiley, 1966.

Atkinson, J. W., & Raynor, J. O. (Eds.). Motivation and achievement. Washington, D.C.: Winston, 1974.

Bentler, P. M., & Bonett, D. G. Significance tests and goodness of fit in the analysis of covariance structures. Psychological Bulletin, 1980, 88(3), 588-606.

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

Cronbach, L. J. Coefficient alpha and the internal structure of tests. Psychometrika, 1951, 16(3), 297-334.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. Psychological Bulletin, 1955, 52(4), 281-302.

Entwisle, D. R. To dispel fantasies about fantasy-based measures of achievement motivation. Psychological Bulletin, 1972, 77(6), 377-391.

Fineman, S. The achievement motive construct and its measurement: where are we now? British Journal of Psychology, 1977, 68, 1-22.

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

Joreskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, and P. D. Suppes (Eds.), Contemporary developments in mathematical psychology (Vol. 2). San Francisco: W. H. Freeman, 1974.

Joreskog, K. G. Structural analysis of covariance and correlation matrices. Psychometrika, 1978, 43, 443-477.

Joreskog, K. G. Addendum to "A general approach to confirmatory maximum likelihood factor analysis". In K. G. Joreskog and D. Sorbom (Eds.), Advances in factor analysis and structural equation models. Cambridge, Mass.: Abt Books, 1979.

Joreskog, K. G., & Sorbom, D. COFAMM user's guide. Chicago: National Educational Resources, 1976.

Klinger, E. Fantasy need achievement as a motivational construct. Psychological Bulletin, 1966, 66(4), 291-308.

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

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. The achievement motive. New York: Irvington, 1953.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. A scoring manual for the achievement motive. In J. W. Atkinson (Ed.), Motives in fantasy, action, and society. Princeton, N. J.: Van Nostrand, 1958.

McClelland, D. C., & Winter, D. G. Motivating economic achievement. New York: Free Press, 1969.

Murray, H. A. Explorations in personality. New York: Wiley, 1938.

Murstein, B. I. Theory and research in projective techniques. New York: Wiley, 1963.

Reuman, D. A. Ipsative behavioral variability and the quality of thematic apperceptive measurement of the achievement motive. Journal of Personality and Social Psychology, in press.

Rokeach, M. The nature of human values. New York: Free Press, 1973.

Veroff, J., Depner, C., Kulka, R., & Douvan, E. A comparison of American motives: 1957 vs. 1976. Journal of Personality and Social Psychology, 1980, 39(6), 1249-1262.

Veroff, J., Douvan, E., & Kulka, R. A. The inner American: A self-portrait from 1957 to 1976. New York: Basic Books, 1981.

Veroff, J., & Feld, S. Marriage and work in America: A study of motives and roles. New York: Van Nostrand Reinhold, 1970.

Wheaton, B., Muthen, B., Alwin, D. F., & Summers, G. F. Assessing reliability and stability in panel models. In D. R. Heise (Ed.), Sociological Methodology 1977. San Francisco: Jossey-Bass, 1977.

Footnotes

[1]We believe an analysis of the variance-covariance matrix is warranted over an analysis of the correlation matrix for two reasons. First, it is logically impossible to test the fit of an appropriate null model for the parallel measures model, when a matrix of correlations serves as the empirical basis for the test. Such a null model hypothesizes that all item variance is random error variance and that the magnitude of this error variance is equal for all items. The variances of all items have necessarily been constrained to be equal in a correlation matrix, however, so this equality constraint cannot be examined when a correlation matrix is used. For models where this type of constraint is not present, it does not matter whether one analyzes a correlation or covariance matrix. The results, except for scale, are identical. Second, as long as the indicators of a construct use the same metric, we believe the item variances are inherently informative and should not be standardized.

[2]To specify a null model that is appropriate for (i.e., more restrictive than) the parallel measures modal, we must constrain error variances to be equal for all elements in the diagonal of Φ that correspond to indicators of a single construct. If a parallel measures model cannot be meaningfully specified in principle, as is the case when the indicators of a construct use different response metrics, an appropriate null model does not constrain error variances to be equal.

[3]Since an interviewer may have inaccurately transcribed a TAT story, and the story would have been coded both for the achievement motive and for story length, we hypothesized that some correlated error may be present for an indicator of motive strength and a corresponding

indicator of story length which were based on the same TAT story (Campbell and Fiske, 1959; Alwin, 1974). Even though the correlated error assumption is somewhat compelling on conceptual grounds, we found no significant improvement in fit for several models which included this assumption, in addition to the assumptions of Models 2 through 6.

[4]The slightly larger reliability estimate derived from the congeneric measures model in Figure 2, compared to that derived from the tau-equivalent measures model in Figure 1, is consistent with expectations based on classical test theory.

[5]As in the previous section, we found that the assumption of correlated error variance does not significantly improve the fit of models in Table 3.

[6]Since the indicators of work satisfaction do not use the same metric, measurement models more restrictive than the congeneric measures model are inappropriate. The parallel measures and the tau-equivalent measures models, which assume equal true score variance in all indicators of a construct, are only plausible if the indicators have at least been measured on the same scale.

Table 1

Fit Indices for Measurement Models

of 1976 Male TAT n Achievement

Model	X^2	df	X^2/df	p-value
1. Null model	81.477	9	9.053	<.0001
2. Parallel measures model	41.392	8	5.174	<.0001
3. Tau-equivalent measures model	4.552	5	.910	.4730
4. Congeneric measures model	.760	2	.380	.6838

Table 2

Fit Indices for Models of 1976 Male

n Achievement, Story Length, and Education

Model			X^2	df	X^2/df	p-value
1.	Null model		1558.274	42	37.102	<.0001
	Measurement Models for					
	n Achievement	Story Length				
2.	C	C	27.367	26	1.052	.3903
3.	T	C	41.396	29	1.427	.0636
4.	C	T	31.438	29	1.084	.3451
5.	T	T	45.694	32	1.428	.0553
6.	C	P	44.098	32	1.378	.0755

Note. C = congeneric measures; T = tau-equivalent measures;
P = parallel measures.

Table 3

Fit Indices for Models of 1976 Male

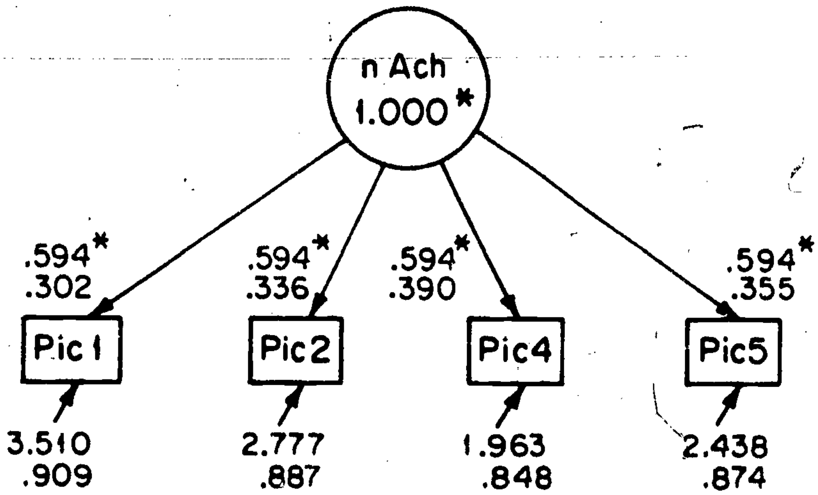
n Achievement, Story Length, and Work Satisfaction

Model				X^2	df	X^2/df	p-value
1.	Null model			1890.711	84	22.508	<.0001
	Measurement Models for						
		Story	Work				
	n Ach	Length	Satisfaction				
2.	C	C	C	85.288	62	1.376	.0266
3.	T	C	C	98.601	65	1.517	.0045
4.	C	T	C	88.543	65	1.362	.0278
5.	T	T	C	102.054	68	1.501	.0047
6.	C	P	C	98.809	68	1.453	.0087

Note. C = congeneric measures; T = tau-equivalent measures;

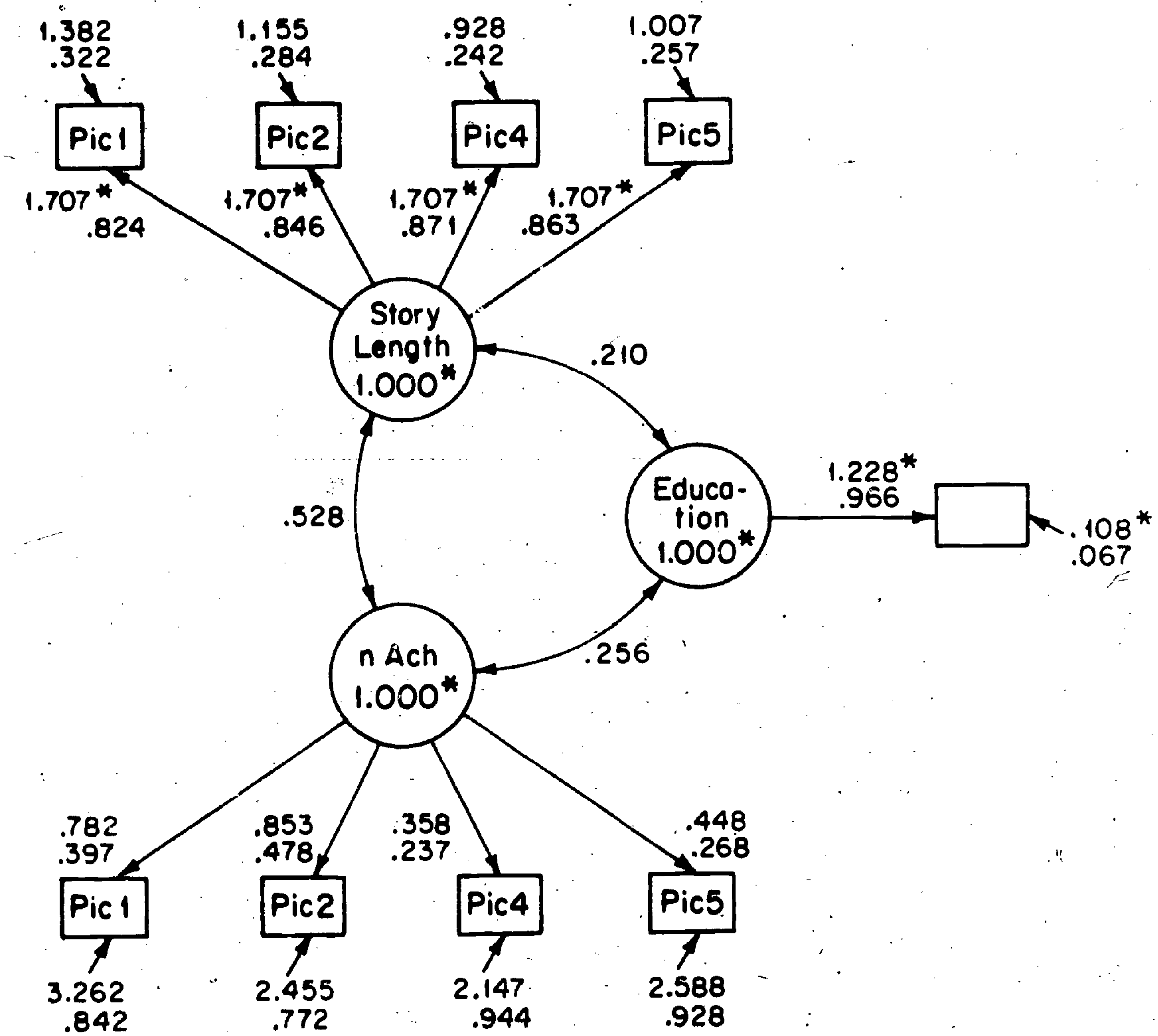
P = parallel measures.

Figure 1
Preferred Measurement Model for 1976 Male TAT n Achievement



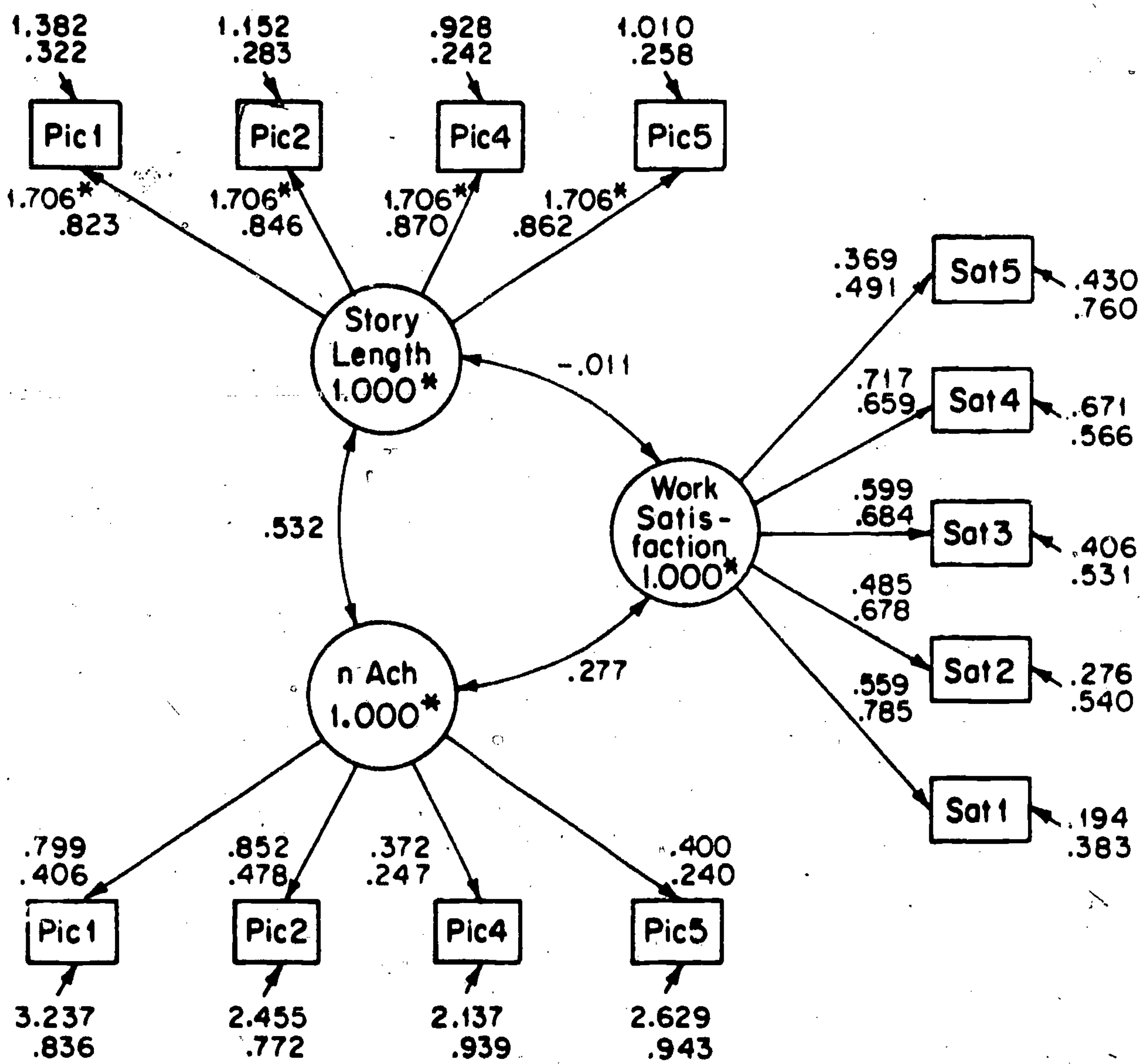
Note. Upper coefficients are COFAMM parameter estimates when only the factor is standardized. Lower coefficients are standardized with respect to the variance of the indicators as well. Starred values were set or constrained to be equal in the initial solution.

Figure 2
Preferred 3-Factor Model for Representing the Covariances
Among Indicators of TAT n Achievement, TAT Story Length, and Education



Note. Upper coefficients are COFAMM parameter estimates when only the factor is standardized. Lower coefficients are standardized with respect to the variance of the indicators as well. Starred values were set or constrained to be equal in the initial solution.

Figure 3
Preferred 3-Factor Model for Representing the Covariances Among Indicators of TAT n Achievement, TAT Story Length, and Work Satisfaction



Note. Upper coefficients are COFAMM parameter estimates when only the factor is standardized. Lower coefficients are standardized with respect to the variance of the indicators as well. Starred values were set or constrained to be equal in the initial solution.