AN APPLICATION OF THE MULTITRAIT-MULTIMETHOD MATRIX TO THE STUDY OF THE N-ACHIEVEMENT CONSTRUCT

Thesis for the Degree of Ph. D.

MICHIGAN STATE UNIVERSITY

Louis J. Hofmann

1965

This is to certify that the

thesis entitled

AN APPLICATION OF THE MULTITRAIT-MULTIMETHOD MATRIX TO THE STUDY OF THE N-ACHIEVEMENT CONSTRUCT

presented by

Louis J. Hofmann

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling, Personnel Services & Educational Psychology

William W. Farquhar
Major professor

Date February 25, 1966

ABSTRACT

AN APPLICATION OF THE MULTITRAIT-MULTIMETHOD MATRIX TO THE STUDY OF THE N-ACHIEVEMENT CONSTRUCT

by Louis J. Hofmann

This study was an investigation of the construct validity of certain *n*-achievement scales. Nine scales were included in the analysis. Five of the scales were measures of *n*-achievement, while four others were measures of *n*-affiliation. The five scales examined in the study were: 1) the McClelland and Atkinson TAT, 2) the French Test of Insight, 3) the Edwards Personal Preference Schedule, 4) a Self-Report measure of *n*-achievement and *n*-affiliation, and 5) the Michigan State M-Scales.

The construct validity of these scales was assessed by: 1) comparing the intercorrelations of the test scores within a multitrait-multimethod matrix such as that proposed by Campbell and Fiske, 2) examining the correlations of the *n*-achievement and *n*-affiliation scores with intelligence and achievement scores, and 3) a factor analysis of the test scores. The test scores were obtained on a sample of eleventh and twelfth grade male high school students.

The multitrait-multimethod matrix analysis yielded evidence for the convergent validity of the *n*-achievement scales, while the *n*-affiliation scales did not meet this

this requirement. The intercorrelations of the n-achievement scales were generally positive and significantly different from zero. The projective methods were less adequate than the objective methods of measurement.

The correlations of the Edwards Personal Preference Inventory, the Self-Report Inventory, and the M-Scales with Lorge-Thorndike Intelligence scores were positive (r = .286, .167, and .388, respectively), while the Test of Insight and the TAT were related to a lesser degree (r = -.217 and .089, respectively).

The correlations of the motivation scores with academic achievement were similar. The EPPS, the Self-Report Inventory and the M-Scales were positively related (r = .360, .301 and .356), while the TOI and TAT correlations were lower (r = .200 and -.098).

The factor analysis produced results consistent with the Multitrait-Multimethod Analysis. The five n-achievement scales had their highest loadings on the same factor as predicted. The n-affiliation scales had their highest loadings on a second factor.

The failure of projective methods of motivation assessment to demonstrate construct validity in this analysis were attributed to the unreliability of these methods along with response sets unique to the production of thematic material.

AN APPLICATION OF THE MULTITRAIT-MULTIMETHOD MATRIX TO THE STUDY OF THE N-ACHIEVEMENT CONSTRUCT

By

Louis John Hofmann

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Personnel Services,
and Educational Psychology

1965

ACKNOWLEDGEMENTS

The technical assistance and moral support of many persons was necessary in the completion of this thesis. Dr. William Farquhar and Dr. David Krathwohl were especially supportive. I wish also to thank the graduate students at Michigan State University for their helpful comments.

Finally, thanks are also due to my wife Judith for her forbearance during the time required for the completion of the thesis.

TABLE OF CONTENTS

TABLE OF CONTENTS CON'T.

LIST OF TABLES

LIST OF APPENDICES

CHAPTER I

INTRODUCTION TO THE PROBLEM

Farquhar,[1] in his study of motivational factors underlying school achievement, has shown that the M-scales possess what the Technical Recommendations for Psychological Tests and Diagnostic Techniques[2] would call predictive validity. Other studies using the M-scales[3,4] have tended to show concurrent, and "factor validity," but as yet no study has been designed to test what ostensibly was the theoretical background for the construction of the test.

---

[1] W.W. Farquhar, Motivational Factors Related to Academic Achievement, United States Office of Education Cooperative Research Project No. 846, 1963.

[2] American Psychological Association, Committee on Psychological Tests, Technical Recommendations for Psychological Tests and Diagnostic Techniques, Washington, D.C.: APA, 1954.

[3] David A. Payne, "The Concurrent and Predictive Validity of an Objective Measure of Academic Self-Concept," Educational and Psychological Measurement, 1962.

[4] Marion D. Thorpe, "The Factored Dimensions of an Objective Inventory of Academic Motivation Based on Eleventh Grade Male Over- and Under-achievers." Unpublished doctoral dissertation, Michigan State University, 1961.

Many other studies of achievement motivation (*n*-ach)* such as the series conducted by Atkinson,[1] present evidence for validity of *n*-ach tests which ranges from what the Technical Recommendations would call concurrent to what is called construct.

Cronbach and Meehl[2] and Campbell[3] have tended to regard construct validity as perhaps the most important part of test validation efforts, but as yet, the evidence about *n*-ach scales would not indicate that they all possess this desirable quality.

Campbell has indicated that the APA Technical Recommendations regarding content, concurrent, and predictive validity are not sufficient to cover what is generally considered to be exhaustive evidence of test validity is needed because certain interpretations are made of the test scores. As he states:

---

[1]John W. Atkinson, An Introduction to Motivation (D. Van Nostrand and Co., Inc., 1964).

[2]Lee J. Cronbach and Paul E. Meehl, "Construct Validity in Psychological Tests," Psychological Bulletin, Vol. 52, No. 4, July 1955, 281-302.

[3]D.T. Campbell, "Recommendations for APA Test Standards Regarding Construct Trait or Discriminant Validity," American Psychologist, Vol. 15, No. 8, August 1960, 546-553.

*The words achievement motivation, and Need-achievement (*n*-ach) will be used interchangeably throughout this thesis. Affiliation motivation and *n*-aff will be similarly used.

> "In the labels given tests, in statements of intent and descriptive material, many explicit and implicit claims are made," and
>
> "Test constructors and users as we have known them have generally been prone to reifying and hypostatizing, prone to assume that their tests were tapping dispositional syndromes with other symptoms than those utilized in the test."[1]

He believes that the implications of such hypostatizing should be checked, and for this reason the construct validation of tests is necessary.

He distinguishes between two types of construct validity. The first is called <u>trait</u> validity. He says:

> "It is applicable at that level of development still typical of most test development efforts, in which "theory," if any, goes no farther than indicating a hypothetical syndrome, trait, or personality dimension."[2]

The second could be called <u>nomological</u> validity and would

> "represent . . . the possibility of validating tests by using the scores from a test as interpretations of a certain term in a formal theoretical network, and through this, to generate predictions which would be validating if confirmed when interpreted as still other operations and scores."[3]

For example, if the n-ach scores are shown to be correlated with teachers' ratings of "academic achievement needs," then trait validity is shown.* If the n-ach

---

[1] <u>Ibid</u>., p. 546.

[2] <u>Ibid</u>., p. 547.

[3] <u>Ibid</u>., p. 547.

*Trait-validity is demonstrated for both measures, the n-ach scale and teachers' ratings. As Campbell states "validation is symmetrical and equalitarian. The presumptive validity of both tests is increased by agreement." p. 548

scores are interpreted as a measure of Drive (D) in the Hullian theory of learning, and correct predictions are made regarding performance in learning situations, then nomological validity would be demonstrated.

It is clear, however, that nomological validity may be too great a requirement at this time for all n-ach measures, and that about all which could be expected is evidence of trait validity.

Need for the Study

Clarification of Constructs

It is generally held that the development of theory, the testing of the theory, and the redevelopment of theory is an efficient approach to science. An opposing view held by some others is that science is essentially a description of the relationships among events and that many empirical laws must be found before an overall integration of some of these empirical laws is attempted.

Most practicing researchers probably operate between these two modes of attack and develop their theories from empirical facts which are not exhaustive of the phenomena in question, and revamp their theories as new research investigates other phenomena encompassed within the constructed theory.

The question of constructs (concepts) in science has been of interest to many philosophers of science, but there is no general agreement as to how one gives meaning

to these constructs. One method has been to define operationally the construct and limit the "meaning" of the construct to the specified set of operations. A strict adherence to this rule would indicate that the "meaning" of intelligence for any investigator is the score on a particular IQ test. Hempel[1] specifies that other properties of the concept may be spoken of which may not be directly related to the operations of the operational definition, and the empirical relationships between the concept (in this case some measure or score) and other concepts.

For example, the concept of Drive (D) in the Hullian framework may be operationally defined by the number of hours of food deprivation, or the voltage of current on a grid, and these measures might be related to the speed or latency of rat's journey down a runway. Once developed, however, the concept of D may have other properties imputed to it. It could be also thought of as a general tendency to react to cues directly related to the deprivation, and to be unreactive to all others, e.g., to "notice" cues for food when a deprivation schedule for food has been set up, and not to "notice" cues for water, if a deprivation of

---

[1]C.G. Hempel, "Fundamentals of Concept Formation in Empirical Science," International Encyclopedia of Unified Science, Vol. II, No. 7, Chicago: University of Chicago Press, 1952.

water has not been part of the experiment. These other properties of D would require further experimentation in order to give D a wider meaning than just a tissue deficit related to food, or a general activation related to noxious stimulation of the footpads by electricity.

These considerations lead to the question of concepts and constructs in psychological theory. The construct of need or motive has had a long history in psychology. It has been usually used in the wider sense (not operationally defined) in many theories of personality. Some theorists like Murray[1] provide an operational definition of need (scores yielded by a special scoring on the Thematic Apperception Test), other theorists use the terms need and motive in a more loose sense and provide only a verbal network of relationships between needs.

There are other issues involved in the meaning of the concepts of need or motive which may have to be resolved, but these are for the most part the province of the theorists who use the term in divergent senses. It would appear, however, that the two terms, need and motive, are used in much the same way by most researchers.

Because the concept of achievement need, and/or achievement motive has recently been the focal point of

---

[1]H.A. Murray, Explorations in Personality (New York Oxford Press, 1938).

much theoretical and empirical work,[1,2,3] it is important that the concept and its measurement be subjected to further empirical test.

In Farquhar's research, the concept of achievement motivation has been related to a criterion of academic achievement, while McClelland and Atkinson have related their broader concept to recall of tasks, self ratings, academic achievement, production records and economic behavior of societies among other things. As new relationships are found, the concept takes on a broader meaning.

It is at this point that the question of construct validity comes in. The question becomes whether measures of *n*-ach differentially fit the predicted theoretical relationships between need scores and achievement behavior. A second question is whether all the tests of achievement measure an underlying trait, and are thus different from other motivation measures.

Various objections to and clarifications of construct validity have appeared. The issues discussed regarding construct validity are generally those of terminology and philosophy of science rather than exhortations

---

[1]Farquhar, op. cit.

[2]D. McClelland, J. Atkinson, et al., The Achievement Motive (New York: Appleton-Century-Crofts, 1953).

[3]A.L. Edwards, Personal Preference Schedule Manual (The Psychological Corporation, 1954).

to eliminate requirements subsumed under the topic of construct validity. Some writers would abandon the notion of construct validity,[1,2] while others would add or change the specific requirements for construct validation. However, a number of attributes of construct validity remain:

a. In tests that presumably measure a trait, syndrome or construct, evidence for the presence of these attributes cannot be gathered from the test alone, but must come from relationships with other tests and non-test behavior.

b. Test scores should acquire meaning from psychological theory (Loevinger[3]; Cronbach and Meehl[4]), although the level of theory development may vary tremendously. Some of the statements in the theory should lead to predicted relationships among observables.

c. Construct validation is both convergent and divergent, that is, relationships between test scores and other variables should be high in some cases and low or non-existent in others.

d. Constructs may vary from being very close to observation to ones highly removed from observation (intervening variables or hypothetical constructs).

---

[1]Robert L. Ebel, "Must All Tests Be Valid?," The American Psychologist, 1961, 640-647.

[2]Harold Bechtoldt, "Construct Validity: A Critique," American Psychologist, Vol. 14, No. 10, October, 1959, 619-629.

[3]Jane Loevinger, "Psychological Tests as Instruments of Psychological Theory," Psychological Reports, 3, 1957, 635-694.

[4]Cronbach and Meehl, op. cit.

Some of the writers consider construct validity as the Sine qua non of validity while others stress predictive and content validity as sufficient. The single issue pervading most of the discussion is that of inferences which we make from test scores. This is evident in the following examples: (a) Cronbach and Meehl's[1] discussion of the nomological net, in which tests scores are used as variables scores which are used as terms in the theoretical net; (b) Campbell's[2] notion of trait and nomological validity in which test scores are seen as measures of a trait possessed in varying degrees by different individuals; (c) Ebel's[3] substitution of meaningfulness for validity in which test score relationships to other measures is an important aspect of meaningfulness; and (d) Loevinger's[4] notion of "structural and external component" in which relations of items to themselves and to total scores and non-test behavior are considered evidence of construct validity.

---

[1] Ibid.

[2] D.T. Campbell and D.W. Fiske, "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, Vol. 56, No. 2, 1959, 81-105.

[3] Ebel, op. cit.

[4] Loevinger, op. cit.

Problem of Validation of Tests

Whenever a new test is used in a research study, a usual "validational" procedure encountered is the correlation of this new test with a whole host of other tests or other "criterion" variables. Occasionally, the terms "construct validity," "predictive validity," are employed as justification for the employment of this new test.

Intelligence tests are validated by other intelligence tests, scholastic aptitude tests are validated by achievement in high schools and colleges, personality tests are validated by efficient discrimination of groups of persons presumably variant on these personality traits, and so forth. Validation techniques are almost as numerous as the number of tests and are potentially more numerous.

What then constitutes adequate validation of psychological tests?

Every test constructor is aware that his test should possess the qualities of validity and reliability, reliability being presumed to be a necessary predecessor of validity. There are types of reliability as well as validity and the propriety of each is determined within the framework of the use of the test.

Once reliability of the test is presumed adequate, the efforts of the test constructor are focused on validity

considerations. Since the APA Technical Recommendations[1] were published, a neat classification of validity has been available to test constructors. Four types of "validity" were established with the necessity of each dependent upon the type of test and the varieties of inferences which were to be made from each test score.

In general, test scores are not seen as important in their own right when construct validity is being considered. Test scores usually represent fundamental entities which are measured in some fashion by a particular test. When content, concurrent and predictive validity are assessed, the test score may be the datum of importance rather than any entity or trait or syndrome.

The problem of this study is basically that of establishing what evidence is available regarding the construct validity of measures of achievement motivation. Two aspects of the problem could be isolated. First, whether the "construct" of achievement motivation is a valid one, and second, whether measures of achievement motivation have "construct" validity. On the face of it, these questions would appear to be similar; however, different methodologies would be employed to answer the separate aspects.

---

[1]American Psychological Association, Committee on Psychological Tests, op. cit.

In the first case a theoretical network could be erected in which the term "achievement motive" was used as a term in the axioms or theorems of the theory. If this theory generated empirical relations which were confirmed, the construct of "achievement motivation" would also be validated. This conceivably could occur even when separate or distinct measures or tests of achievement motivation were used as operational definitions in the research. However, one measure of "achievement motivation" could be more efficient than others in the various predictions made by the theory. In this case, one measure possesses greater construct validity than the others and the question of the validity of the construct is assessed concomitantly with the assessment of the construct validity of the measure of achievement motivation.

The case, however, is not nearly as neat when single tests are considered when they are based on less developed theory or on the hunches and intuitions of the test constructor. When one measures introversion, n-ach, self concept, ego, nurturance or any of a whole host of traits, which have more or less of a place in theoretical or nomological nets, the problem is more difficult. Perhaps a number of independent measures relate and are called "achievement motivation" measures. Do these relationships signify the construct validity of the test or the validity of the construct (achievement motivation)? It would appear

that evidence for the validity of a theory is also evidence for the constructs embedded in the theory and is mutually confirming of the construct validity of the tests used as measures of these constructs. It is difficult to imagine cases in which "construct validation" of tests in the sense of Campbell and Cronbach and Meehl could proceed without explicit theoretical or empirical statements and as such, construct validation of tests probably coincides with validation of constructs. (For opposing view, see Kausler and Trapp[1].)

This study does not attempt to solve the philosophical problems attendant to the use of concepts such as "construct validity." It does, however, extract some minimal relevant requirements for construct validity and apply these to measures of achievement motivation.

New Empirical Evidence

There are at least three other vacuums that research on achievement motivation could attempt to fill.

The first vacuum is caused by the fact that most research on achievement motivation has focused on college students. The attenuation of empirical relationships

---

[1]D.H. Kausler and E.P. Trapp, "Methodological Considerations in the Construct Validation of Drive Oriented Scales," Psychological Bulletin, 56, 1959, 152-157.

exhibited by the studies reviewed in Chapter II is potentially related to the relative *n*-ach homogeneity of the samples studied. There is also the second vacuum, caused by the small number of variables used in the research. A slightly larger scale study will provide a more extensive empirical study of the relationships among measures of achievement motivations.

Finally, the explicit job of differential prediction is a void in the liturature. Few studies have checked the differential association of achievement motivation to the variables of aptitude and achievement.

Summary of Need for Study

The need of this study is assumed to proceed from:

a. Lack of clarity of the concept of achievement motivation.

b. Lack of validity information concerning *n*-ach measures.

c. The lack of evidence concerning the empirical relationship of achievement motivation to other variables, especially at the high school level.

The Purpose of the Study

The first and main purpose of the study is to provide evidence about convergent and discriminant validity of various achievement motivation scales. The purpose will be therefore to provide evidence for the trait validity of achievement motivation scales.

2. A second purpose is to extend the information available regarding the predictive validity of achievement motivation scales. Academic achievement behavior, along with aptitude test responses, will be considered as criterion measures.
3. A third purpose is to explicitly submit motivational scales to the job of prediction. Motivational measures will be compared concerning their ability to predict achievement behavior.

Theory

The theory which underlies this research study is the need theory of Murray[1] and its refinements and redevelopments by McClelland, et al.,[2] Atkinson[3;4;5] and Farquhar.[6]

---

[1]Murray, op. cit.

[2]McClelland and Atkinson, et al., op. cit.

[3]Atkinson, op. cit., An Introduction . . . .

[4]John W. Atkinson (ed.), Motives in Fantasy Action and Society, D. Van Nostrand Co., Inc., 1958.

[5]John W. Atkinson, et al., "The Achievement Motive, Goal Setting and Probability Preferences," Journal of Abnormal and Social Psychology, 60, No. 1, 1960, 27-36.

[6]Farquhar, op. cit.

Murray conceived of needs as constructs which stood for forces in the brain region which organize apperception and actions to transform unsatisfying situations in a certain direction. Murray believed that these needs could be aroused from within by internal visceral processes, or by the immediate situation. These needs were manifested by typical behaviors, avoidance and selection of certain cathected objects, characteristic affect, manifestation of satisfaction with the achievement of a certain effect, and dissatisfaction caused by the failure to achieve these effects.

The Thematic Apperception Test was designed to assess these underlying needs. Murray and others devised systems for scoring the content of these stories. Early studies of the TAT demonstrated that fear and hunger affected the imagery produced on the TAT. These studies led to the conception of the relationship between motivation and projective responses.

McClelland and Atkinson[1] developed group methods of administering the TAT and standardized the questions which guided the production of responses by the subject. Analysis of stories given under food deprivation gave support to the belief that thematic apperceptive content was susceptible to motivational influence. Studies of the TAT

---

[1]McClelland and Atkinson, et al., op. cit.

under various conditions -- relaxed, neutral, success, failure, and others -- demonstrated that achievement imagery was related to the type of induced motivational condition. It was found that stories following achievement orientation contained references to (a) performing a task in relation to a standard of excellence; (b) performance of a task socially defined as a unique accomplishment; and (c) pursuit of a long-term goal, the characters of the stories being concerned with "success".

These studies provided the foundation for a series of studies which ultimately lead to the formalized conception outlined by Atkinson.[1] Although most of McClelland's and Atkinson's investigations focused on the achievement motive, the affiliation motive has also been shown to be related to behavior. Performance under "non-ego involved" conditions was positively correlated with affiliation motive scores, and was unrelated to n-ach level.

These modifications and elaborations of Murray's original need theory lead to tests used in this study. The TAT method and its scoring is outlined in Atkinson.[2] The

---

[1]Atkinson, op. cit., An Introduction . . . .

[2]Atkinson, op. cit., Motives in Fantasy . . . .

Test of Insight (TOI) is described by French.[1;2] The Edwards Personal Preference Schedule,[3] which is scored for affiliation and achievement need, has also a similar theoretical rationale.

Farquhar[4] conceptualized academic motivation as being a bi-polar dimensional construct which considered low motivation as the polar opposites of McClelland's three characteristics of achievement motivation. Thus, (a) meeting a minimal standard of excellence, (b) performing tasks which are considered common accomplishment and (c) having short-term goals, became the characteristics describing persons with low academic motivation. Academic motivation became merely a special case of achievement motivation in an academic setting.

Atkinson's recent statements[5] place the original concept of achievement motivation into a more specific form. He states that the tendency to achieve success (T_s)

---

[1]Elizabeth G. French and Irene Chadwick. "Some Characteristics of Affiliation Motivation," Journal of Abnormal and Social Psychology, Vol. 52, No. 3, 1956, 296-300.

[2]Elizabeth G. French, "Some Characteristics of Achievement Motivation," Journal of Experimental Psychology, Vol. 50, No. 4, 1955, 232-236.

[3]Edwards, op. cit.

[4]Farquhar, op. cit.

[5]Atkinson, op. cit., An Introduction . . . .

("overtly expressed in the direction, magnitude, and persistence of achievement and oriented performance") is a product of the joint action of the motive to achieve success (M_s), the strength of the expectancy or probability of success (P_s), and the incentive value of success (I_s).

The equation $T_s = M_s \times P_s \times I_s$ describes the interrelationships of these variables. M_s is conceived to be a relatively stable and general characteristic of the person while P_s and I_s are specifically determined by the experience a person has had in similar performance situations.

Atkinson also posits a general motive to avoid failure, this motive being measured by the Test Anxiety Questionnaire and the Manifest Anxiety Scale. The theory underlying these uses of the tests is different than that underlying the development of these anxiety questionnaires.

Some relevant implications for this study emerge from Atkinson's theorizing. He states[1] that T_s should be greatest when the P_s is at .50, and should be lower if P_s is very high or low. He also states when the apparent difficulty of the tasks is held constant for a group, then the T_s should be greater when M_s is strong than when M_s is weak. This difference again should be most pronounced when P_s is near .50.

---

[1] Ibid.

Also, when there is no incentive to achieve, there will be no basis to predict the performance of persons differing on M_s. These implications are important for this study because some of the experimental comparisons evaluate the relative relationship of achievement motivation scores and affiliation motive scores to academic performance. A positive correlation between achievement motivation and academic performance is reasonable because

a. Some tasks in school are likely to have near .50 difficulty -- that is, not all tasks are extremely easy or impossibly difficult, so M_s (n-ach) is likely to be operative in academic behavior.

b. P_s (perceived probability of success) is probably unrelated to achievement motivation (M_s) so analysis of individual probability (P_s) for each subject is not necessary.

c. Most schools have incentives for academic performance; e.g., grades and academic honors. These incentives may, however, be extremely idiosyncratic to each school, but it would seem to be reasonable to assume that many academic tasks have incentive value for most students.

If it can be assumed that there is no interaction between P_s and I_s and M_s in the samples studied, then a straightforward prediction about performance in indicated.

The theory underlying affiliation motivation is very similar to that of achievement motivation. No formal schema has been provided for it, but some hypotheses regarding its action in performance situation were demonstrated by French and Chadwick.[1]

Its place in the study is mainly one of control. Other traits could have been chosen, except that three previously investigated instruments yield scores on this variable.

The main technique of the present study, the multitrait-multimethod matrix, was outlined by Campbell and Fiske.[2] Its main purpose is to demonstrate that, given a valid test, correlations between certain variables should be high, and correlations between other variables should be low. This matrix led to the major hypotheses of the study, although the variables chosen for inclusion in the study were chosen by the theory outlined above.

The Hypotheses of the Study

1. Independent measures of the N-ach trait will correlate positively, and significantly different from zero.

---

[1]French and Chadwick, op. cit.

[2]Campbell and Fiske, op. cit.

2. The correlation of two independent measures of the *n*-ach trait, will be larger than any correlation of these trait measures and any other trait measured by the same or different methods.
3. Independent measures of the same trait will be more highly correlated than either of these independent measures will be correlated with other traits measured by the same method.
4. The correlations between the self-report achievement motivation *n*-ach scores and academic achievement scores will be lower than the other correlations between achievement motivation scores and academic achievement scores.
5. The correlations between achievement motivation scores and academic achievement scores will be higher than the correlation between affiliation motivation scores and academic achievement scores with method being the same.
6. The correlation between achievement motivation scores and intelligence scores will be lower than the correlation between achievement motivation scores and academic achievement scores with methods controlled.
7. The correlation between affiliation motivation trait measures and intelligence will be lower than the correlation of affiliation motivation and academic achievement.

8. The factor loadings of the n-achievement tests will be concentrated on a first order factor, with lower loadings on a second factor, a factor which will have high loadings of n-affiliation tests.

Overview of the Study

An analysis of the problem of construct validity is made in the first chapter. The problem of construct validity as it relates to n-achievement measures is similarly treated in the first chapter.

The second chapter contains a review of the literature dealing with construct validity, and also contains a review of theory and empirical characteristics of tests of achievement motivation.

The design and the analysis of the study are presented in chapters three and four. The fifth chapter contains the summary and conclusions of the study.

CHAPTER II

REVIEW OF THE LITERATURE

The literature deemed relevant to this study would appear to separate into four areas. The first section is a review of some of the historical background of the construct validity criterion; a second is some examples of putative research of test construct validity; a third section contains reviews of studies of the reliability of achievement motivation scales; and a fourth area reviews validity studies of achievement motivation scales. It is not intended to be an exhaustive overview of research of achievement motivation, but is designed to place construct validity in a historical perspective, and show results of representative studies using measures of achievement motivation, especially as they relate to academic achievement behavior.

Types of Validity

In the history of psychological testing, only recently have exhortations been made by people concerned with test construction that test publishers should show evidence of construct validity when claims of test validity are made.

One of the earliest expressions of this concern was the APA Manual, "Technical Recommendations for Psychological Tests and Diagnostic Techniques".[1] Contained in this publication is one of the first elaborations of the term, construct validity. Its definition of construct validity is best understood when compared to its definitions of the other types of validity.

Content Validity

The manual states

> "Content validity is evaluated by showing how well the content of the test samples the class of situations or subject matter about which conclusions are to be drawn." (p. 13)

In general, decisions about content validity are made by judging the correspondence between the test behavior and the behavior which the test presumably samples. Decisions about many educational tests are more easily made than are decisions regarding unstructured personality tests (such as the Rorschach or TAT), and in the latter case, one makes some assumptions about the relationships between the two sets of behaviors (e.g., in the TAT the stories told in the pictures are assumed to be derived from unconscious needs, but some relationship has to be postulated between verbal reports and underlying dynamics).

---

[1] American Psychological Association, Committee on Psychological Tests, op. cit.

In work samples, arithmetic and vocabulary tests, content validity is more easily evaluated since the test behavior is in some cases replicas of the situations to which the conclusions are made.

Predictive Validity

> "Predictive validity is evaluated by showing how well predictions made from the test are confirmed by evidence gathered at some subsequent time. The most common means of checking predictive validity is correlating test scores with a subsequent criterion measure." (p. 13)[1]

When "validity" coefficients are usually reported they are of the predictive validity sort, or the concurrent validity sort which will be explained below.

Concurrent Validity

> "Concurrent validity is evaluated by showing how well test scores correspond to measures of concurrent criterion performance or status. Studies which determine whether a test discriminates between presently identifiable groups are concerned with concurrent validity." (p. 14)[2]

The predictive validity notion and the first aspect of the concurrent validity notion are practically identical. The decision between whether one or the other is being exhibited would depend on whether the two behaviors are seen as synonymous or not. Presumably, when the test performance precedes criterion performance, predictive validity

---

[1] Ibid., p. 13.

[2] Ibid., p. 14.

is shown, and concurrent validity is shown when both measures are taken at the same time.

The second aspect of concurrent validity seems to be slightly different from the first -- namely, the discrimination between presently identifiable groups, or the correspondence between current status and test performance. This raises an interesting question regarding the usual conception of concurrent validity. Since inclusion in identifiably different groups is presumably based on prior behavior (schizophrenics and manic depressives exhibited their symptomatology before diagnosis and inclusion), and the discrimination is ostensibly about presently constituted groups, it, nonetheless, really has "predicted" behavior prior to the test behavior. Another example might be the "prediction" of cumulative grade point averages, which in some cases are constituted by behaviors prior to the testing period. The Technical Recommendations have no category for such "predictions" since "predictions after the fact" are usually an anathema to the scientific investigator. Another category, however, could be evaluated which could be called "postdictive validity," which includes aspects and considerations such as the above.

Construct Validity

The Technical Recommendations state

> "Construct validity is evaluated by investigating what psychological qualities a test

> measures, i.e., by demonstrating that certain explanatory constructs account to some degree for performance on the test. To examine construct validity requires both logical and empirical attack. Essentially, in studies of construct validity, we are evaluating the theory underlying the test. The validation procedure involves two steps. First, the investigator inquires: From this theory, what predictions would we make regarding the variation of scores from person to person or occasion to occasion? Second, he gathers data to confirm these predictions." (p. 14)[1]

Following closely the statement of the Technical Recommendations, Cronbach and Meehl attempted to clarify the term construct validity. A construct, they said, "is some postulated attribute of people, assumed to be reflected in test performance. In test validation the attribute about which we make statements in interpreting the test is a construct."[2]

According to Cronbach and Meehl, these attributes have occasionally inadequate criteria and their meaning is inherent in the network of associations or propositions in which the construct occurs. Validation of the construct or construct validation occurs only when the statements in the network lead to predicted relationships among observable events. They go on to state that construct validity is examined under many types of evidence including:

---

[1] Ibid.

[2] Cronbach and Meehl, op. cit.

a. content validity

b. inter-item correlations

c. inter-test correlations

d. test-criterion correlation

e. stability

f. stability under experimental intervention

Excessively high correlations and stability may constitute confirming or negative (disaffirming) evidence. When predicted relationships do not occur, the fault is either in the network of laws or in the construct. Cronbach and Meehl also conclude that construct validity cannot generally be expressed in the form of a single coefficient. Variance attributable to the construct can however be roughly estimated. They see construct validation as not essentially different from the scientific enterprise of theory development and confirmation.

The essence of Cronbach and Meehl's definition is its requirement that the construct be systematically investigated in theories, and predicted relationships are the evidence for the acceptance of a construct. Jessor and Hammond[1] made similar recommendations using manifest anxiety as a measure of D (drive) in Hull's mathematical learning theory.

---

[1]R. Jessor and K.R. Hammond, "Construct Validity and the Taylor Anxiety Scale," Psychological Bulletin, Vol. 54, 1957, pp. 161-170.

At about the same time, Loevinger[1] attempted to show how test construction itself is a procedure which has important relationships to construct validation. She distinguishes three types of validity:

a. substantive

b. structural

c. external

Her substantive category is closely akin to the content validity and external validity encompasses much of what is included within predictive and concurrent categories of the technical recommendations. Of construct validity, she says,

> "The basic concept is that of the construct validity of the test, the degree to which it measures some trait which really exists in some sense. Construct validity can only be established by convergence of several lines of evidence. Evidence for construct validity can be broken down into evidence that the test measures something systematic, and evidence for the particular interpretation of what it measures."[2]

Loevinger concurs with the requirements of the Technical Recommendations and seems to add some unique facets. Her structural validity category includes evidence similar to the Technical Recommendations and adds factoral structure and homogeneity of items as new evidence. The content and reliability of the test are also stressed more

---

[1] Loevinger, op. cit., p. 685.

[2] Ibid.

than in the Technical Recommendations or in Cronbach and Meehl.[1] Although she is highly interested in the nature of items and their interrelationships, she says, "External correlations provide the court of last appeal."[2]

Loevinger's[3] requirements for validity are sweeping and extensive. She raises an issue which deserves further treatment, i.e., whether evidence for the validity of a test is evidence for the validity of the construct. This issue is similarly raised by Kausler and Trapp insofar as manifest anxiety and "D" are concerned. They maintain that both processes cannot coexist. In their example, the use of the Manifest Anxiety Scale (MAS) was criticized as a measure of D, and that validation of the construct D is more adequate if experimentally induced D states were used as a measure of D. A fallacy here is that experimentally induced D states are no more of a final criterion for D than is the Manifest Anxiety Scale. In the second experiment using experimentally induced drive states as a measure of D, they are providing another line of evidence for the validity of the construct D, and for the measure (experimentally induced states). In the experiment, using MAS as a measure of D, both the construct D and the MAS were being validated. The

---

[1]Cronbach and Meehl, op. cit.

[2]Loevinger, op. cit., p. 675.

[3]Ibid., p. 659.

issue raised by Kausler and Trapp is really whether experimentally induced positive drive states (or some other operational definition of D) or MAS stand in a more preeminent position as a measure of D.

Recently other criticisms and clarifications of construct validity have emerged. Bechtoldt[1] criticized the concept of construct validity because the category allowed private interpretations of test scores and behavior and lead to unnecessary confusion attendant to the concept. He proposed a return to an "operational methodology" which regarded test scores as operational definitions. He states:

> "What is being questioned is the tendency to consider as part of public empirical science hunches involving 'vague' ill defined variables and relations between such variables, these hunches being derived primarily from the observed performance."[2]

He goes on to criticize the circularity inherent in construct elaboration which uses the construct to explain performance after deriving the construct from observed performance.

Campbell[3] answered many of Bechtoldt's objections and proposed new distinctions. He states that the predictive and concurrent aspects might be better called "practical validity." These occur primarily when there

---

[1]Bechtoldt, op. cit.

[2]Ibid.

[3]Campbell, op. cit.

are "criteria" against which the tests might be judged. Campbell[1] points out, however, that "not all psychological tests have been designed solely to predict performance against extant institutional decisions, situations. There are relatively few settings which produce such criteria." The scientist is however more interested in single factored traits "for which society produces no correspondingly pure criteria."

He pointed out that "Cronbach and Meehl, Jessor and Hammond, have tended to tie construct validity to tests developed and validated in the context of explicit theoretical structures or 'nomological nets.'"[2] Such developed theory is usually lacking in test validation efforts. He would therefore divide construct validity into two types.

> Trait Validity "is applicable at that level of development still typical of most test development efforts in which 'theory,' if any, goes no further than indicating a hypothetical syndrome, trait, or personality dimension."[3]
>
> Nomological Validity "would represent the very important and novel emphasis of Cronbach and Meehl on the possibility of validating tests

---

[1] Ibid., p. 547.

[2] Ibid., p. 627.

[3] Ibid., p. 547.

> by using the scores from the tests as interpretations of a certain term in a formal theoretical network and through this to generate predictions which would be validating if confirmed when interpreted as still other operations and scores."[1]

Somewhat earlier Campbell and Fiske[2] advocated the use of the multitrait-multimethod matrix as a method of test validation. Their method provides evidence for convergent and discriminant validity. (Both incidentally are also predictive, concurrent, practical and trait validity previously described.)

Discriminant validity is demonstrated when correlations between independent methods measuring different traits are consistent with prior knowledge concerning the relationship between these different traits. The relationships may be positive or negative, but should in any case be quite low.

Ebel[3] was also concerned with the notion of validity. He points out correctly that validity has a good deal of confusion attendant to its use in the literature. This leads to a suggestion that a word "meaningfulness" be substituted for validity and the requirements for meaningfulness are outlined. It is not clear to this reviewer how these

---

[1] Ibid.

[2] Campbell and Fiske, op. cit.

[3] Ebel, op. cit.

requirements differ from those under the separate rubric of content, concurrent, predictive and construct validity.

Summary of Construct Validity

In some ways, validity now means very little more than what Cureton[1] proposed earlier. New terms have supplanted and been superimposed on older terms, but operational requirements for validity are substantially similar to those of previous years. Correlations of test scores with other scores constitute the main method of test validation with logical and statistical analysis of test content serving a subsidiary role. The theoretical efforts preceding the computation of the correlations, the types of scores which are correlated, and the interpretation of the coefficients are the main differentiators of the different types of validity.

Recently, other validational techniques have come increasingly into vogue. Factor analysis is one prominent example. Guttman's facet analysis is another. For the most part, a factor analysis of a test or a battery of tests gives similar information as the MT-MM Matrix, although prior factor structure is not usually hypothesized as are the interrelationships of the MM-MT Matrix. Guttman's facet analysis probably functions both as a test generator

---

[1]Edward Cureton, "Validity," in E.F. Lindquist, Educational Measurement, ACE, Washington, 1951, 621-692.

and construct validator. Its heavy emphasis on the use of theory to generate tests would seem to be highly desirable. Even with the general use of factor techniques, there have been few, if any, rationalizations of this method into the hypothetico-deductive methods of science, and it remains an exploratory (inductive) method of concept development. Guilford and Eysench (see Cronbach and Meehl, reference 16) present factor analysis as a method of test and construct validation. And Cronbach and Meehl also support the use of factor analysis in construct validity although it does not fit their more restrictive requirements. In general, these methods are used in early stages of construct development.

Examples of Construct Validation Studies in the Psychological Literature

A few studies selected from the psychological literature were reviewed in order to determine the variety and scope of investigations which purport to analyze the construct validity of tests, and/or the validity of constructs.

A study of Phares and Adams[1] examined the validity of the EPPS heterosexuality scale. Extreme groups of males on the Heterosexuality Scales ranked "sexual" and "nonsexual" pictures which had been previously equated esthetically.

---

[1]E.J. Phares and C. Adams, "The Construct Validity of the Edwards PPS Heterosexuality Scale," Journal of Consulting Psychology, 25, 1961, 341-344.

The high scorers on the Heterosexuality Scale ranked sexual pictures higher than the low scorers. A high group, when exposed to sexual communication, learned and retained the material more easily than a low group. The second hypotheses received statistical support however only after eliminating subjects.

Bernardin and Jessor[1] used two EPPS scales to define "dependent" and "independent" subjects. The scales used were the Autonomy and Deference scales. Percentile separations on the two scales of 30 points with 70th and 50th percentiles as cut-off points was the mode of classification. Three hypotheses were examined. The hypothesis that dependent persons would perform a finger maze mask less efficiently under negative verbal reinforcement, and the hypothesis that dependent persons will seek more help in a problematical situation (when assistance is available on request) were confirmed. However, a third hypothesis that dependent groups would conform more often to group judgments than would independent groups was not supported.

Zuckerman, et al.,[2] factor analyzed various measurements and ratings of dependency for 72 student nurses. They

---

[1]A.C. Bernardin and R. Jessor, "A Construct Validation of the Edward's Personal Preference Schedule with Respect to Dependency, Journal of Consulting Psychology, 21, 1957, 63-67.

[2]M. Zuckerman, et al., "Concurrent and Construct Validity of Direct and Indirect Measuring Dependency," Journal of Consulting Psychology, Vol. 25, No. 4, 1961, 316-323.

found low relationships between projective measures of dependency and self and peer ratings. However, the EPPS and a sentence completion test were positively related to self and peer ratings. The factor analysis was judged by the authors to be the relevant analysis of "construct validity."

The first two studies indicate that the construct validity of tests is demonstrated when high and low scorers on a particular trait behave appropriately to what is known about the trait. The theory is the trait description. The second study[1] used two scale scores to define a third trait called dependency. The confirmation of hyptheses regarding dependency presumably lent evidence for the construct validity of the original measures (autonomy and deference).

The third study[2] assessed construct validity with factor analysis, with the exact validity evidence being unspecified.

Apparently, then, construct validation procedures do not have a common base, and perhaps all of the usual validity evidence is also construct validity evidence (a position similar to that of Loevinger[3]).

---

[1]Bernardin and Jessor, op. cit.

[2]Zuckerman, et al., op. cit.

[3]Loevinger, op. cit.

Two studies which explore the validity problems similarly to that of the MT-MM matrix are those of Dicken[1] and Hills[2]. The first study assesses the validity of the California Psychological Inventory. Two methods (the CPI and Staff Ratings) were used to measure five traits (dominance, responsibility, impulsiveness, intellectual competence, and rigidity). The convergent criterion was met satisfactorily by 4 of 5 CPI variables. However, only 2 CPI variables met the discriminant criterion, i.e., certain CPI variables correlated too highly with other variables. Hills' study while not purporting to examine convergent and discriminant validity, does provide the raw material for such a matrix. See Table 1. He attempted to modify the general notion of McClelland regarding n-ach into four kinds of achievement motivation. Specially selected pictures and questionnaires were designed to tap these four kinds of achievement motivation. There were economic, social, academic, and professional. These methods are labelled I and II respectively. Method III are two other sets of pictures previously used to measure achievement motivation.

---

[1]Charles F. Dicken, "Convergent and Discriminant Validity of the California Psychological Inventory," Educational and Psychological Measurement, 23, 3, 1963, 449-459.

[2]John R. Hills, "Needs for Achievement, Aspirations, and College Criteria," Journal of Educational Psychology, 41, 1, 1958, 156-161.

TABLE 2.1

A MODIFIED MULTI-TRAIT--MULTI-METHOD MATRIX USING DATA FROM HILLS' STUDY OF ACHIEVEMENT NEEDS AND ASPIRATION LEVELS

		METHOD I				METHOD II				METHOD III	
		A_1	B_1	C_1	D_1	A_2	B_2	C_2	D_2	E	F
Method I	A_1		-.08	.16	.15	-.02	-.04	.06	.02	-.02	.11
	B_1			.32	.39	.16	.32	.06	.28	.35	.47
Picture	C_1				.14	.11	.04	-.16	.06	.25	.13
	D_1					.04	-.04	-.04	-.02	.14	.34
Method II	A_2					.83	.51	.35	.33	.19	.00
	B_2						.81	.46	.60	.09	.01
Questionnaire	C_2							.69	.72	.02	.04
	D_2								.80	-.03	.20
Method III	E										.33

The resultant picture is one of poor convergent validity and fair discriminant validity. The highest correlations were located in the mono-method-heterotrait block for the questionnaire method. Most of the other high correlations were in the mono-method-heterotrait block of the picture method. This suggests strong method variance with very small shared trait variance. This could be due to the separation of the n-ach score into four components, and the unreliability of the picture method.

Reliability of Achievement Motivation Instruments

In the following section, reliabilities of instruments used to measure achievement motivation which were reported in the literature are reviewed. The review will be restricted to three of the commonly used tests (TAT, EPPS, and TOI).

The most widely used test was the TAT with the McClelland scoring system. There were some studies reporting the reliability of the TOI while few studies analyzed the reliability of the EPPS achievement scale.

Most of the studies of the TAT seemed content to report interscorer reliabilities and they were generally quite high. Karolchuk and Worell[1] found an interscorer

---

[1]P. Karolchuk and L. Worell, "Achievement Motivation and Learning," Journal of Abnormal and Social Psychology, 53, 1956, 255-257.

agreement in a high school student sample of 108. Marlowe[1] found a Pearson r of .96 in a sample of 44 undergraduate males. Reitman[2] found an interscorer r of .90 with 108 college students and Burdick[3] reported an interscorer r of .90 for a large army sample. Similar results were found by Melikian.[4] However, somewhat lower interrater r's were found by Himelstein, et al.,[5] in an Air Force Academy sample of 298 males. They reported an interscorer reliability of .714 before a scoring conference.

---

[1]D. Marlowe, "Relationships Among Direct and Indirect Measures of the Achievement Motive and Overt Behavior," Journal of Consulting Psychology, Vol. 23, No. 4, 1959, 329-332.

[2]Walter R. Reitman, "Need Achievement, Fear of Failure, and Selective Recall," Journal of Abnormal and Social Psychology, Vol. 62, No. 1, 1961, 142-144.

[3]Harry Burdick, "Interrelationships of Three Measures of Motivation," Psychological Reports, 8, 1961, 225-226.

[4]Levon H. Melikian, "The Relationships Between Edward's and McClelland's Measures of Achievement Motivation," Journal of Consulting Psychology, Vol. 22, No. 4, 1958, 296-299.

[5]P. Himelstein, A.E. Eschenbach, and A. Carp, "Interrelationships Among Three Measures of Need Achievement," Journal of Consulting Psychology, Vol. 22, No. 6, 1958, 451-452.

Three studies considered stability of achievement motivation scores. Birney[1] in a study of two forms of the TAT found "equivalence-stability" coefficients of r = .03 to r = + .56. He considered the five-month interval stability coefficient of r = + .29 as representative of six coefficients reported. Krumboltz and Farquhar[2] reported test-retest coefficients of r = .49, .02, .27, .25 for various subgroups of undergraduate college students. Their total group had a test-retest r = .26, which is about the median of all coefficients reported. However, their inter-scorer reliability estimate was r = .91. Kagan and Moss[3] investigated the stability of achievement motivation scores over a two-year nine month and five-year nine month period. The stability is reported in phi-coefficients using the presence or absence of achievement fantasy as the raw data. The phi-coefficients were +.16, +.22 and +.32. They also reported a 95% "agreement" between two raters. One study analyzed the internal consistency of the TAT achievement

---

[1]Robert C. Birney, "The Reliability of the Achievement Motive," Journal of Abnormal Social Psychology, 58, 1959, 266-267.

[2]J.D. Krumboltz and W.W. Farquhar, "Reliability and Validity of the -ach Test," Journal of Consulting Psychology, Vol. 21, No. 3, 1957, 226-228.

[3]Jerome Kagan and Howard A. Moss, "Stability and Validity of Achievement Fantasy," Journal of Abnormal and Social Psychology, Vol. 58, No. 3, May, 1959.

motivation test. Hills[1] found an internal consistency estimate of r = .64 on his sample of 56 law school students. This was the average inter-item correlation for 18 pictures, corrected by the Spearman-Brown formula. Another analysis indicated that average inter-item r's on his four subtests was r = .18, .02, .23, .10. In view of these low relationships it is surprising to find his total test reliability estimate so high.

Four studies which used the TOI had reliability information; all four being studies of interscorer reliability. French[2] reported interscorer reliability of r = .88 and .91 for successive samples of 90 officer candidates. Atkinson, et al.,[3] found an interscorer reliability of r = .90 on a sample of 66 college males. Himelstein, et al.,[4] found an interscorer reliability of .70, which is very similar to the coefficient they found for the TAT n-achievement scoring. Van Zandt and Himelstein[5] reported both alternate form and

---

[1]Hills, op. cit.

[2]French, op. cit.

[3]Atkinson, et al., op. cit., "The Achievement Motive . . . .

[4]Himelstein, Eschenbach, and Carp, op. cit.

[5]B.R. Van Zandt and P. Himelstein, "The Role of Verban Fluency on a Projective Measure of Motivation," Educational and Psychological Measurements, Vol. 19, No. 4, 1959, 663-666.

test-retest reliabilities for both achievement motivation and affiliation motivation. These were r = .24 and r = .37 for achievement motivation, and r = .08 and r = .70 for affiliation motivation, respectively. Partialing out verbal fluency did not appear to change these estimates. "Rescore" reliabilities of r = .86 and r = .90 were also reported respectively for achievement motivation and achievement affiliation scores.

Himelstein and Kimbrough[1] checked the test-retest and alternate form reliability of the Test of Insight. While their test-retest coefficients were very low (r = .36 total group) the alternate form reliability was even lower (r = .17).

Uhlinger and Stephens[2] administered the EPPS to 72 college freshmen and found the internal consistency reliability estimate to be r = .79. This was the highest of all internal consistency estimates reported.

Summary of Reliability Studies

The reliability of n-ach measures appears to present certain difficulties in interpretation. On the one hand,

---

[1]R. Himelstein and W.W. Kimbrough Jr., "Reliability of French's 'Test of Insight'," Educational and Psychological Measurement, Vol. 20, No. 60, 1960, 737-741.

[2]Uhlinger and Stephens, op. cit.

it is quite apparent that scoring formulas lead to high interscorer agreement. With only one dissenting finding, these coefficients are all in the .90's; on the other hand, stability and internal consistency reliability estimates are all quite low. In light of these findings, it is no mystery why researchers have tended to limit their reliability information to the inter- and intrascorer type. There would appear to be no detectable difference between the TOI and the TAT in the interscorer reliability. One study reporting need-affiliation reliability would indicate that this trait is somewhat more stable, although interscorer reliability is no better than that of achievement motivation.

Validity Studies of Achievement Motivation Tests

Interrelationships Among Tests of Achievement Motivation

A number of studies have analyzed the interrelationship of the EPPS, TAT, TOI. Bendig[1] found the TAT and the EPPS to be uncorrelated (r = .11). Melikian[2] found similarly that the EPPS and the TAT had a low relationship (r = +.16). But since the test was administered to 84 Arab students, he concluded that the TAT may be affected by the

---

[1]A.W. Bendig, "Manifest Anxiety and Projective and Objective Measures of Need Achievement Items," Journal of Psychological Studies, 11, 1959, 32-38.

[2]Melikian, op. cit.

culture. Marlowe[1] found a negative relationship between the two achievement motivation measures. However, his r of -.05 was determined on a sample of only 44 undergraduate students.

Weiss, et al.,[2] in a sample of 60 college males found that the TAT and the EPPS were moderately related, r = .26. Mitchell,[3] in a study of 131 nursery students, found that the relationship between the TAT and other specially constructed motivation instruments was slightly negative although these more objective instruments did have some variance in common (r = .40, .13, .35).

Himelstein, et al.,[4] intercorrelated the TOI, the EPPS, and the TAT in their study of 298 members of the freshman class at the Air Force Academy. These intercorrelations hovered around zero, for both raters. His interscorer reliability was, however, the lowest of the studies of reliability cited above.

---

[1]Marlowe, op. cit.

[2]P. Weiss, M. Wertheimer and Byron Groesbeck, "Achievement Motivation, Academic Aptitudes, and College Grades," Educational and Psychological Measurement, Vol. 19, No. 4, 1959, 663-666.

[3]James J. Mitchell, "An Analysis of the Factorial Dimensions of the Achievement Motivation Construct P," Journal of Educational Psychology, 1961, 179-187.

[4]Himelstein, Eschenbach and Carp, op. cit.

Shaw,[1] however, found the TOI and the EPPS to be moderately related r = .51, .26 for the male samples. However, these two scores were negatively related in the female samples (r = -.15 and r = -.13). The TOI and the TAT had low relationships in both the male and female samples (r = .25 and r = .09 and r = .28 and r = .20, respectively). These relationships could be due to the small sample sizes and the breakdown of the sample into "achiever" and "underachiever" subsets. His groups also had unusual homogeneity in intelligence, especially for a high school group. An IQ of above 110 was necessary for inclusion in the study.

It is clear that the interrelationships among these achievement motivation is very moderate. An upper bound of r = .26 seems reasonable, while a lower bound is zero. This occurs even though interscorer reliabilities on these instruments is usually quite high.

Achievement Motivation and Intelligence

Krumboltz and Farquhar[2] found the TAT and the ACE psychological exam to be unrelated. The coefficients within their various subgroups varied around zero. A specially

---

[1]M.C. Shaw, "Need Achievement Scales as Predictors of Academic Success," Journal of Educational Psychology, Vol. 52, No. 6, 1961, 282-285.

[2]Krumboltz and Farquhar, op. cit.

constructed aptitude measure devised by Weiss, et al.,[1] also had low relationships with both the TAT and the EPPS. The respective r's of .10 and .18 emerging from their sample of 60.

Two other studies supply information about the relationship between aptitude scores and motivation scores. Mitchell[2] found the TAT-ACE relationship to be low (r = .11), Bendig[3] similarly found both the TAT and EPPS to have zero relationships with a verbal aptitude measure (r = .02, r = .03, respectively). Longnecker[4] reported, however, a correlation of r = .23 between ACE scores and EPPS n-ach scores (N = 292 college sophomores.

Achievement Motivation and Academic Performance

Several investigations were concerned with the correlation of achievement motivation and various academic achievement measures, at both the college and high school levels. These studies are summarized in Table 2.

---

[1]Weiss, Wertheimer and Groesbeck, op. cit.

[2]Mitchell, op. cit.

[3]Bendig, op. cit.

[4]E.D. Longnecker, "Perceptual Recognition As a Function of Anxiety, Motivation and the Testing Situation," Journal of Abnormal and Social Psychology, Vol. 64, 1961, 215-221.

TABLE 2.2

SUMMARY OF STUDIES USING ACHIEVEMENT MOTIVATION SCORES TO PREDICT ACADEMIC PERFORMANCD

Investigator	N	Level	Test	Results
Bendig[1]	263	College	EPPS Items	r = -.09 --- +.10
Uhlinger & Stephens[2]	72	College	EPPS "	"no significant relationship"
Weiss, et al.[3]	60	College	EPPS "	r = .42
Jordan & DeCharms[4]	60	High School	TAT "	r = .05
Hills[5]	56	College	TAT "	r = -.22
Krumboltz & Farquhar[6]	169	College	TAT "	r = .13; .12
Wiss, et al.	60	College	TAT "	r = .34
Mitchell[7]	131	College	TAT "	r = .08
Riccuiti 1954[8]			TAT "	r = .16; .29
Herron[9]	180	College	TOI "	r = .05

[1]Bendig, op. cit.

[2]Uhlinger and Stephens, op. cit.

[3]Weiss, Wertheimer and Groesbeck, op. cit.

[4]T.E. Jordan and R. DeCharms, "The Achievement Motive in Normal and Mentally Retarded Children," American Journal of Mental Deficiency, 64, 1959, 457-466.

[5]Hills, op. cit.

[6]Krumboltz and Farquhar, op. cit.

[7]Mitchell, op. cit.

[8]Henry N. Riccuiti, "The Prediction of Academic Grades with a Projective Measure

It is clear that if a relationship exists between achievement motivation scores and academic achievement, it is quite low. McClelland's[1] reported correlation using the TAT and Weiss's EPPS achievement correlation is the only relationship of any significance. The few negative correlations were probably due to restriction of range; Hills[2] study was on law school grades, although Jordan's and DeCharms'[3] had a normal high school sample. The reported coefficient was only slightly negative (r = .05), however.

Other studies of the relationship between achievement motivation and academic achievement produced conflicting results. Littig and Yeracaris[4] found achievement motivation related to achievement for men but not for women (n = 190, 206). In Shaw's[5] study only the TOI was related

---

of Achievement Motivation: Initial Validating Studies," Princeton, New Jersey, ETS, 1954.

[9]E. Wayne Herron, "Intellectual Achievement Motivation: A Study in Construct Clarification," The University of Wisconsin, Milwaukee, 1964.

[1]McClelland, Atkinson, et al., op. cit.

[2]Hills, op. cit.

[3]Jordan and DeCharms, op. cit.

[4]Lawrence W. Littig and Constantine A. Yeracaris, "Academic Achievement Correlates of Achievement and Affiliation Motivation," Journal of Psychology, 55, No. 1, 1963, 115-119.

[5]Shaw, op. cit.

to the achiever-underachiever classification. The TAT and the EPPS were in the predicted direction, but the differences were not statistically significantly different. Two other studies reported negative relationships between achievement motivation and achievement behavior. Cole, et al.,[1] divided their studies into over- and under-achiever samples using students who fell 1 S.D. above or below the regression line. This produced 8 over- and 13 underachievers. A Wilcoxen matched pairs test produced a statistic with a .06 probability, indicating differences higher in achievement motivation (TAT) in the underachieving sample. They attributed the differences between McClelland's and their findings to be the non-life-like situation in the laboratory studies.

Somewhat earlier, Broverman, et al.,[2] found that high "strivers" produced significantly more achievement fantasy themes than did low strivers. The size of their sample was rather small (N = 34) for the confidence of their proposals. They contended that their study supports the "alternative channels" interpretation of the expression of

---

[1]D. Cole, et al., "The Relation of Achievement Imagery Scores to Academic Performance," Journal of Abnormal and Social Psychology, Vol. 65, No. 3, 1962, 208-211.

[2]Donald M. Broverman, Edward J. Jordan and Leslie Phillips, "Achievement Motivation in Fantasy and Behavior," Journal of Abnormal and Social Psychology, 60, No. 3, 1960, 374-378.

achievement motivation, which states that achievement fantasy may serve as a substitute channel for the expression of achievement motivation when this expression is blocked.

Summary of Ach Motivation - Academic Performance Studies

In summary, achievement motivation seems to be quite unrelated to the academic performance criterion. Two criticisms of the studies appear to be in order. For the most part the samples studied were very small. *N*'s below 100 are more likely than not. Secondly, most of the studies have employed homogeneous groups of subjects. Nearly all of the studies reviewed were college samples and some within narrow professional areas. Had studies been conducted which had less restriction in the range of ability and achievement, greater relationships may have been obtained.

Validity Studies of the M-Scales

Four studies were conducted using the M-Scales to predict academic achievement. Green,[1] in a study of 104 male and 129 female selected high school students, found rather high correlations between the subtests of the M-Scales and Grade Point Average. The M-Scale-aptitude test score correlation was also extremely high.

---

[1]Robert Lee Green, "The Predictive Efficiency and Factored Dimensions of the Michigan State M-Scales for Eleventh Grade Negro Students, An Exploratory Study," Thesis for the Degree of Ph.D., Michigan State University, 1962.

Kipfmueller[1] also found high correlations between the sub-tests of the M-Scales and Grade Point Average for a sample of 200 randomly selected parochial high school students. The correlations of the M-Scales to an aptitude measure were substantial, but generally lower than the M-Scale-GPA correlation. For males, however, the correlations between HIT and PTCS and GPA were lower than HTI, PTCS and an aptitude score. Johnson[2] found lower correlations generally between all subtests and GPA and aptitude. HTI and PTCS also were more highly correlated with aptitude than with GPA. In published data by Farquhar,[3] similar correlations between HTI and PTCS and both aptitude and GPA were obtained.

Total M-Scale scores were found to correlate highly with GPA by Farquhar (r = .40, 48 females; r = .56, r = .49 males), but the correlations between M-Scales total scores and APT are similarly quite high (r = .30, r = .43 females;

---

[1]Mark K. Kipfmueller, "The Predictability and Factored Dimensions of the M-Scales for Eleventh Grade Parochial School Students," unpublished doctoral dissertation, Michigan State University, 1963.

[2]Van C. Johnson, "An Assessment of the Motivation Factor in the Estimation of Academic Achievement of Eleventh Grade Indian Students and the Factored Dimensions of the M-Scales. An Exploratory Study. An unpublished doctoral dissertation, Michigan State University, 1963.

[3]Farquhar, op. cit.

r = .45, r = .50 males). So, in at least one of the comparisons, the correlation of motivation scales and aptitude is higher than the correlation of motivation scales and achievement behavior, a situation clearly opposite to the theory surrounding the three trait measures.

The reason for these high mutual relationships probably stems from the method of test development, in which discrimination between "over- and under-achievers" was the criterion for item selection. Had over- and under-achievers been matched on aptitude such high correlations with aptitude would probably not have emerged.

Unpublished data collected under the direction of Farquhar indicated that in another sample (n = 179 males, 189 females) the correlation between M-Scales and an aptitude score was r = .25, r = .45 for males and females respectively, while the correlation between M-Scales and GPA was r = .52 and r = .54 for these same groups; this relationship fitting more closely the theory regarding achievement motivation, aptitudes and achievement behavior.

CHAPTER III

DESIGN OF THE STUDY

Sample

The theoretical population of individuals from which the sample was selected was 11th and 12th grade high school male students. After contacting a number of schools (five) three agreed to participate in the study. The following samples resulted from these contacts.

Sample A: 22 11th grade male students from a medium sized town.

Sample B: 30 11th grade male students from a rural consolidated school district.

Sample C: 60 students (all 11th and 12th grade males) from a small town.

The major portion of the research problem did not seem to indicate that representatives of sample was a particular requisite. What was necessary was a population with sufficient heterogeneity so that the trait measures would have sufficient variance.

Instrumentation

One standardized commercial test was used (EPPS) while the others, although being used for research studies, were not commercial. The M-Scales, TAT, TOI and a specially constructed self-report inventory were also administered.

The M-Scales -- The battery is composed of four subtests. They are:

The Generalized Situational Choice Inventory. A forced choice instrument which elicites choices between situational events which were designed to compare the two polar dimensions of academic motivation.

The Preferred Job Characteristics Scale. This is also a forced choice instrument with types of jobs being the stimulus material.

The Word Rating List. This scale is composed of a series of adjectives to which the subject responds as to their perception of teachers' beliefs about them. The range of applicability is from "always" to "never."

The Human Trait Inventory. This scale is made of a series of statements describing various human behaviors. The subject again rates these behaviors on a four point scale as to their applicability to him (or her).

The Thematic Apperception Test (TAT) -- Six TAT-type pictures were used in the present study. These pictures were made from slides purchased from the American Documentation Institute and are pictures of Atkinson's list.[1] The verbal

---

[1]Atkinson (ed.), op. cit., Motives in Fantasy . . . .

Number	
4	Cub reporter scene
25	Boy watching large airliner
26	Foreman and worker standing near machine
33	Boy smiling at desk at home
88	A group of youngmen seated, with one young man standing
97	A chemist in a white coat

productions made to the pictures are assumed to contain expressions of need for achievement and need-affiliation. They are scored according to Atkinson's[1] scoring manual.

The French Test of Insight (TOI) -- The TOI is a series of ten statements about human behavior. The subject's task is to explain the behavior. Twenty minutes is allowed for the completion of the ten items. The test is scored similarly to the TAT, and also yeilds both a *n*-ach and a *n*-aff score. Both the TOI and the TAT were scored by James Mullin, an advanced doctoral student in Psychology.

The Edward's Personal Preference Schedule (EPPS) -- The EPPS is a series of items in which the person chooses one of two paired statements. Each of the pair is a statement referring to one of 15 needs. Permission was secured from the Psychological Corporation to reproduce 56 of the

---

[1] Ibid.

items,* all of which include either affiliation achievement need items, or both. It is an ipsative scale, and high scores on one trait necessarily imply somewhat lower than maximum possible on all other traits. Of the 56 items, four couple achievement and affiliation items.

This ipsativity was partially reduced by eliminating two items on each scale which had statements of both needs.

The TAT, TOI, and EPPS yield both affiliation and motivation scores, while the M-Scales yields only a motivation score.

One other *n*-ach measure was also collected for the study. A self-rating inventory contained a rephrased description of affiliation and achievement needs and asked the subjects to rate themselves from 0 - 100 on these trait descriptions. This scale is reproduced in the Appendix. Three other measurements were collected on Sample C. 1963 and 1964 grades, Lorge-Thorndike intelligence tests and retests of the M-Scales after a nine-week interval were supplied for students in this sample.

---

*Since there is some evidence that item responses obtained to selected items isolated from the context of the Edwards Personal Preference Schedule are not comparable to those obtained within the context, the results of this research cannot be considered applicable to the standardized complete form of the EPPS.

Statistical Hypotheses

For clarity in the following discussion, the variables contained in the hypotheses will be numbered.

1. TAT *n*-achievement score
2. TAT *n*-affiliation score
3. TOI *n*-achievement score
4. TOI *n*-affiliation score
5. EPPS *n*-achievement score
6. EPPS *n*-affiliation score
7. Self-report *n*-achievement score
8. Self-report *n*-affiliation score
9. M-Scales Total score
10. GPA - 1963 average
11. GPA - 1964 average
12. Lorge-Thorndike IQ score

The following hypotheses proceed from a general assumption that, if a trait measure has construct validity, then certain empirical relationships between that trait measure and other trait measures obtain. A research procedure which erects null hypotheses and confirms the research hypotheses by rejecting these null hypotheses is not clearly appropriate for all the research hypotheses.

A special note about the purpose of the study may be helpful in interpretation of the hypotheses. The study's main intent is to study the validity of the achievement motivation construct. As such, no one measure of achievement motivation occupies a preeminent position as the criterion measure. All the tests are presumptive measures of the same trait. (While all the test developers may not consider this to be true, it is curious that they all build their theoretical rationale from Murray and McClelland.) Thus, a correlation exhibited within the matrix may provide

mutual evidence for the validity of two tests or the invalidity of both tests. The question of what is evidence for that is complicated by the fact that: (1) one particular trait measure may have a longer and more glorious history than others; and (2) the matrix considered independently of prior research provides both evidence for the validity of the construct, and the construct validity of the trait measures. The question really being, is the trait measure in question, or is the construct in question. The possible outcomes of this study could shed light on both issues.

There is also a question about the role of affiliation motivation. While it is not the central focus, the research bears similarly on its construct validity. It was chosen as a contrasting or comparable trait because some of the trait measures of achievement motivation also yielded scores on this trait. All the evidence considered together provides the basis for the judgment of construct validity of trait measures and validity of the construct. Rejection of the null hypothesis may not be sufficient evidence for this judgment; therefore, the following hypotheses are stated in the directional form.

Statement of Hypotheses

H: 1. Independent measures of the n-ach trait will correlate positively, and significantly different from zero.

A. Symbolically

H: -achievement $r_{13}, {}^{4}_{35}, r_{57}, {}^{4}_{79}, r_{15}, r_{37}, r_{59}, r_{17}, r_{39}, r_{19} > 0$

and

H: -affiliation $r_{24}, r_{46}, r_{68}, r_{26}, r_{48}, r_{28} > 0$

H: 2. The correlation of two independent measures of the -ach trait, will be larger than any correlation of these trait measures and any other trait measured by the same or different mentods.

Symbolically

H:

A $r_{13} > r_{\underline{14}}, r_{16}, r_{18}, r_{\underline{23}}, r_{36}, r_{38}$

B $r_{35} > r_{\underline{45}}, r_{25}, r_{\underline{36}}, r_{38}, r_{23}, r_{58}$

C $r_{57} > r_{47}, r_{27}, r_{\underline{67}}, r_{\underline{58}}, r_{25}, r_{45}$

D $r_{79} > r_{69}, r_{49}, r_{29}, r_{\underline{89}}, r_{27}, r_{47}, r_{67}$

E $r_{15} > r_{14}, r_{18}, r_{45}, r_{\underline{16}}, r_{\underline{25}}, r_{58}$

F $r_{37} > r_{\underline{47}}, r_{67}, r_{27}, r_{36}, r_{\underline{38}}, r_{34}$

G $r_{59} > r_{89}, r_{\underline{69}}, r_{49}, r_{29}, r_{58}, r_{45}, r_{25}$

H $r_{17} > r_{14}, r_{16}, r_{18}, r_{27}, r_{47}, r_{67}$

I $r_{39} > r_{38}, r_{36}, r_{\underline{29}}, r_{49}, r_{69}, r_{89}, r_{32}$

J $r_{19} > r_{\underline{29}}, r_{49}, {}^{4}_{69}, {}^{r}_{89}, r_{18}, r_{16}, r_{1}$

H: 3. Independent measures of the same trait will be more highly correlated than either of these independent measures will be correlated with other traits measured by the same method.

Symbolically

1. $r_{13} > r_{12}$ and r_{34}
2. $r_{15} > r_{12}$; r_{56}
3. $r_{17} > r_{12}$; r_{78}
4. $r_{19} > r_{12}$
5. $r_{35} > r_{34}$, r_{56}
6. $r_{37} > r_{34}$, r_{78}
7. $r_{39} > r_{34}$
8. $r_{57} > r_{56}$; r_{78}
9. $r_{59} > r_{56}$
10. $r_{79} > r_{78}$

H: 4. The correlations between the self-report -ach scores and academic achievement scores will be lower than the other correlations between achievement motivation scores, measures and academic achievement scores.

$r_{7\ 11} < r_{3\ 11}; r_{5\ 11}; r_{9\ 11}$ and $r_{1\ 11}$

H: 5. The correlations between achievement motivation scores and academic achievement scores will be higher than the correlation between affiliation motivation scores and academic achievement scores with method being the same.

$r_{1\ 11} > r_{2\ 11} >$

$r_{3\ 11} > r_{4\ 11} >$

$r_{5\ 11} > r_{6\ 11} >$

$r_{7\ 11} > r_{8\ 11} >$

H: 6. The correlation between achievement motivation scores and intelligence scores will be lower

than the correlation between achievement motivation scores and academic achievement scores with methods controlled.

$$r_{1\ 11} > r_{1\ 12}$$
$$r_{3\ 11} > r_{2\ 12}$$
$$r_{5\ 11} > r_{5\ 12}$$
$$r_{7\ 11} > r_{7\ 12}$$
$$r_{9\ 11} > r_{9\ 12}$$

H: 7. The correlation between affiliation motivation trait measures and intelligence will be lower than the correlation of affiliation motivation and academic achievement.

$$r_{2\ 11} > r_{2\ 12}$$
$$r_{4\ 11} > r_{4\ 12}$$
$$r_{6\ 11} > r_{6\ 12}$$
$$r_{8\ 11} > r_{8\ 12}$$

H: 8. The factor loadings of the -achievement tests will be concentrated on a first order factor, with lower loadings on a second factor, a factor which will have high loadings of n-affiliation tests.

Analysis Procedures

Hypothesis 1 will be tested by comparing the obtained correlations between independent measures of achievement motivation (hereafter known as the validity coefficients), and their sampling distribution given no relationship between these measures.

The statistic used will be

$$t = \frac{r}{\sqrt{1 - r^2}} \sqrt{n - 2}$$

d.f. = n - 2

[1]Paul Blommers and E.F. Lindquist, Elementary Statistical Methods, Houghton Mifflin Co., Boston, 1960, 451

Hypotheses 2 and 3 were examined by counting the reversals in the predicted rankings of the correlation coefficients. The binomial test was applied to the number of predicted relationship with the hypothesis that the probability of a reversal was = .50.

Hypotheses 4 and 5 were tested by examining the differences in the coefficients with a "t" test for the differences between correlation coefficients obtained on the same sample.[1]

---

[1] H.M. Walker and J. Lev, Statistical Inference, Henry Holt & Co., New York, 1953. The formula proposed by Walker and Lev is:

$$t = \frac{r_{12} - r_{13} \quad (n - 3)\ (1 + r_{23})}{)2(1 - r_{12}^2) - r_{13}^2 - r_{23}^2 + 2r_{12}r_{13}r_{23}}$$

CHAPTER IV

ANALYSIS OF RESULTS

Convergent and Discriminant Validity

The first three hypotheses were attempts to examine the convergent and divergent validity requirements outlined by Campbell and Fiske in their discussion of the multi-trait-multimethod matrix. The subjects used were the total membership of samples A, B, and C. Table 4.1 contains the intercorrelations of all the n-achievement and n-affiliation test scores for 112 male high school students. This matrix provides the raw data for the assessment of the first three hypotheses.

The first hypothesis states that the intercorrelations of the n-achievement scales will be positive and will be statistically significant. This is called the convergent validity requirement and is the most basic of all the different validity requirements. It stems from the assumption that traits can be measured in independent ways and that scores yielded by different methods of measurement should relate in a non-chance fashion.

A similar rationale is appropriate for the n-affiliation tests, and date is also presented for the assessment

TABLE 4.1

MEANS, STANDARD DEVIATIONS, AND INTERCORRELATIONS OF THE N-ACHIEVEMENT AND N-AFFILIATION TEST SCORES FOR 112 MALE HIGH SCHOOL STUDENTS

		Variable								
		1	2	3	4	5	6	7	8	9
Method										
A	2	-.289								
B	3	.174	.109							
	4	-.020	-.030	-.056						
C	5	.195	-.156	.168	-.118					
	6	-.014	.167	-.125	.037	-.322				
D	7	.156	-.111	.237	.067	.396	-.170			
	8	.068	.146	.080	.022	-.353	.247	.113		
E	9	.190	-.022	.242	.044	.425	-.074	.597	-.015	

Legend	Variable	Mean	Standard Deviation
A = TAT	1 = N-ach	8.18	6.40
	2 = N-aff	1.57	2.36
B = TOI	3 = N-ach	2.37	2.59
	4 = N-aff	2.36	1.98
C = EPPS	5 = N-ach	13.35	3.77
	6 = N-aff	12.31	3.92
D = Self Report	7 = N-ach	71.00	18.70
	8 = N-aff	79.44	15.53
E = M-Scales	9 = N-ach	96.83	19.97

of convergent validity of n-affiliation scores. Tables 4.2 and 4.3 contain the intercorrelations of the n-achievement and the n-affiliation scores.

The intercorrelations of n-achievement scales are in the predicted positive direction. While being low, seven of the ten coefficients fill the significance requirement. The n-affiliation tests, fail, however, to meet the convergent validity requirement to any degree. While five of the six coefficients are positive, only one is statistically significant.

The second hypothesis states that the correlations of two independent measures of the same trait will be higher than the correlations of these trait measures and other trait measures which have the same or different method of measurement.

This is one aspect of Campbell and Fiske's discriminant validity requirement. "Method" means simply the kind of test, not the trait presumably measured. E.G., the TAT is a method of measurement, as is the EPPS. n-achievement is the trait measured. Together they form a trait-method unit. The hypothesis states that two (independent) methods of measuring the n-achievement trait (two trait-method units with traits in common) will correlate more highly than either of the trait-method units will correlate with other trait-method units (neither trait nor method in common). Thus, the TAT and EPPS methods of measuring n-achievement

TABLE 4.2

INTERCORRELATIONS OF THE TAT, EPPS, TOI, SELF-REPORT INVENTORY, AND M-SCALE N-ACHIEVEMENT SCALES FOR 112 MALE HIGH SCHOOL STUDENTS

	TOI	EPPS	S-R	M-Scales
TAT	.174	.195*	.156	.190*
TOI		.168	.237*	.242*
EPPS			.396*	.425*
S-R				.597*

*Probability r = 0 < .05

TABLE 4.3

INTERCORRELATIONS OF THE TAT, TOI, EPPS AND SELF-REPORT INVENTORY N-AFFILIATION TEST SCORES FOR 112 MALE HIGH SCHOOL STUDENTS

	TOI	EPPS	S-R
TAT	-.030	.167	.146
TOI		.037	.022
EPPS			.247*

*Probability r = 0 < .05

are hypothesized to have a higher correlation than all other correlations of TAT n-achievement and n-affiliation measurements, and all EPPS n-achievement measures and n-affiliation measurements (this does not include the case in which the n-affiliation method is the same as the n-achievement method. This is examined in Hypothesis 3).

Table 4.4 contains all the separate sub-hypotheses for each validity coefficient (the correlations of independent methods of measuring the same trait) with their corresponding values obtained on the sample of 112 students. The asterisks indicate the coefficients which are not in the predicted direction.

All the validity coefficients exceed the hetero-trait, hetero-method coefficients if the sign of the coefficient is ignored. The three coefficients not in the predicted direction are all the same one (the highly negative correlation between EPPS n-achievement and the Self-Report n-affiliation measurements.

The third hypothesis is the more stringent divergent validity requirement. It states that independent measures of n-achievement will be more highly correlated than either of these trait measures will be correlated with other traits measured by the same method. The comparison correlation is the correlation of two traits measured by the same method. In Table 4.5 the hypotheses and the obtained coefficients are presented.

Six of the obtained coefficients are a reversal of the predicted relationship. However, these reversals appear to be the result of strong negative trait intercorrelations when the same method is employed. If direction of the correlation is considered, then all the subhypotheses are confirmed.

TABLE 4.4

COMPARISONS OF N-ACH VALIDITY COEFFICIENTS AND HETEROTRAIT-HETEROMETHOD COEFFICIENTS--THE HYPOTHESES AND OBTAINED COEFFICIENTS

Validity Coefficient		Heterotrait-Heteromethod Coefficients						
r13 .174	>	r14 -.020	r16 -.014	r18 .068	r23 .109	r36 -.125	r38 .080	
r35 .168	>	r45 -.118	r25 -.156	r36 -.125	r38 .080	r23 .109	r58* -.353	
r57 .396	>	r47 .067	r27 -.111	r67 -.170	r58 -.353	r25 -.156	r45 -.118	
r79 .597	>	r69 -.074	r49 .044	r29 -.022	r89 -.015	r27 -.111	r47 .067	r67 -.170
r15 .195	>	r14 -.020	r18 .068	r45 -.118	r16 -.014	r25 -.156	r58* -.353	
r37 .237	>	r47 .067	r67 -.170	r27 -.111	r36 -.125	r38 .080	r34 -.056	
r59 .425	>	r89 -.015	r69 -.074	r49 .044	r29 -.022	r58 -.353	r45 -.118	r25 -.156
r17 .156	>	r14 -.020	r16 -.014	r18 .068	r27 -.111	r47 .067	r67* -.170	
r39 .242	>	r38 .080	r36 -.125	r29 -.022	r49 .044	r69 -.074	r89 -.015	r32 .109
r19 .190	>	r29 .022	r49 .044	r69 -.074	r89 -.015	r18 .068	r16 -.014	r14 -.020

*Probability reversal = .5 < .018

TABLE 4.5

THE COMPARISON OF VALIDITY N-ACH COEFFICIENTS AND HETEROTRAIT-MONOMETHOD COEFFICIENTS SHOWING HYPOTHESES AND OBTAINED VALUES

Validity Coefficient		Comparison Coefficient
H: 1	r13 (.174) >	r12, r34 (-.289, -.056)
H: 2	r15 (.195) >	r12, r56 (-.289, -.322)
H: 3	r17 (.156) >	r12, r78 (-.289, -.113)
H: 4	r19 (.190) >	r12 (-.289)
H: 5	r35 (.168) >	r34, r56 (-.056, -.322)
H: 6	r37 (.237) >	r34, r78 (-.056, .113)
H: 7	r39 (.242) >	r34 (-.056)
H: 8	r57 (.396) >	r56, r78 (-.322, .113)
H: 9	r59 (.425) >	r56 (-.322)
H: 10	r79 (.597) >	r78 (.113)

Legend:

1 = TAT N-ach	6 = EPPS N-aff
2 = TAT N-aff	7 = Self-Report N-ach
3 = TOI N-ach	8 = Self-Report N-aff
4 = TOI N-aff	9 = M-Scales
5 = EPPS N-ach	

If the probability of a correlation being a reverse of one predicted is .5, then the probability associated with the outcome is P = .227 (with signs of coefficients not being considered).

Discussion

The convergent and discriminant validity of n-achievement scores were examined in Hypotheses 1,2, and 3. The

convergent validity requirement is met in a moderate sense, although all of the coefficients are of moderate size. They are, however, all positive and seven of the ten coefficients are statistically higher than zero. The coefficients involving the projective tests are lower than those involving the more objective measurements.

The convergent validity of the *n*-affiliation is clearly inferior to the *n*-achievement tests. Not only are the coefficients lower but one is negative and one is statistically significant. The discriminant validity requirement is also met to a substantial degree. Reversals in the predicted patterns of intercorrelations occur when strong negative relationships between *n*-affiliation and *n*-achievement emerge.

A number of factors may be responsible for the resultant pattern of inter-correlations. The discrimination of the projective tests at this level severely restricts the variance of the motive scores. Many zero scores were recorded for both the TAT *n*-affiliation and the TOI *n*-achievement and *n*-affiliation. This major restriction of variance was a likely factor in the size of the projective test relationships.

Unwanted response sets and methods variance also appears to be present. The negative correlations between *n*-affiliation and *n*-achievement are probably partially a function of methods variance, especially when these traits

are measured by the same method. To what extent a true correlation exists between these two traits is not known. Since there are negative relationships (for example, the EPPS n-ach, Self-Report n-affiliation r = .353) independent of method, it is likely that there is an actual relationship between the two traits.

Methods factors may function differently in the various tests. In the projective tests the elicitation of scorable achievement imagery by the picture may mean that scorable affiliation imagery output is depressed.[1] This would be especially true if timing severely limits the quantity of output. On the other hand, verbal fluency may cause the scorable imagery for both traits to covary. The first explanation would seem to be more appropriate for the TAT.

The forced choice nature of the EPPS with the item overlap increase the likelihood of a spurious negative relationship. The so-called ipsative nature of the EPPS was examined by rescoring the EPPS eliminating two overlapping items (items which matched EPPS n-ach and n-affiliation). The intercorrelation changed from r = .37 to r = -.32. Eliminating all item overlap would likely reduce it further.

---

[1]See R. Christie and F. Lindauer, "Personality Structure," Annual Review of Psychology, Vol. 14, 1963, 217.

The Self-Report method produced a positive relationship between *n*-achievement and *n*-affiliation. The tendency to use the upper or middle parts of the scale may have been a factor in this relationship.

The net effect of these possible response sets and methods factors may have been a factor in the low and negative correlations obtained between the motive measurements. The general tendency would appear to be a minor negative relationship between *n*-achievement and *n*-affiliation variables, with stronger negative relationships appearing when certain of the methods factors operate.

The Predictive Quality of the Motivational Scales

The relationship of the motivational measures to academic achievement and intelligence test performance, and the differential prediction of these criterion measures by the motivational scales are examined in Hypotheses 4 through 7.

Hypothesis 4 states that the correlation of the Self-Report *n*-achievement scores and academic achievement will be lower than the correlations between other achievement scores and academic achievement.

The correlation between Self-Report *n*-achievement scores and 1963 and 1964 GPA was $r = .279$, and $r = .301$. The correlation between 1963 and 1964 GPA was $r = .842$ and can be considered as a stability reliability coefficient. In Table 4.6, the correlations of all the motivation variables

and academic achievement for both 1963 and 1964, the significance tests, and the "t" values are presented.

TABLE 4.6

SIGNIFICANCE TESTS AND CORRELATIONS BETWEEN ACHIEVEMENT MOTIVATION MEASURES AND 1963 AND 1964 GRADE POINT AVERAGES FOR 54 MALE STUDENTS OF SAMPLE C

Motivational Test	GPA 1963	GPA 1964	"t" Comparison of 1964 GPA, with Self-Report
GSCI	.242	.145	
PJCS	.190	.074	
WRL	.397*	.494*	
HTI	.304*	.325*	
M-Scale Total	.375*	.357*	.417
TOI N-ach	.207	.200	-.610
TAT N-ach	-.067	-.098	-.202
Self-Report N-ach	.279*	.301*	
EPPS N-ach	.227	.360*	.418

*Probability r = 0 $<$ .05

The correlation between the Self-Report N-ach measure and 1963 and 1964 grades is nearly as high as all the other N-ach - GPA correlations. The Word Rating List is correlated most highly with GPA with the M-Scale total score and the EPPS exceeding slightly the Self-Report relationship. However, both the TOI, the TAT, the GSCI and the PJCS are only moderately related to academic achievement.

This would tend to indicate that the more elaborate measurement procedure embodied in the TAT and TOI does not

produce a score which relates highly to achievement behavior in any direct fashion. The correlations of the TOI to achievement behavior are about .20 and the TAT is not related at all. This findings is consistent with Mitchell's[1] and Broverman's[2] finding but is inconsistent with other investigators' positive findings[3;4] who found correlations of $r = .34$, and $r = .51$, respectively.

The hypothesis of superiority of complex methods of motivation measurement in the prediction of academic achievement behavior is clearly not supported by these findings. The more complex methods may have a greater ease of interpretation, but they do not correlate significantly higher than the self-rating device with achievement behavior. It can be safely stated, however that to the extent to which one wishes to predict academic grades, a self-rating device is not inferior to more elaborate methods. Certainly other justifications could be made for more complex methods.

The fifth hypothesis states that the relationship between achievement motivation scores and academic achievement will be larger than the relationship between affiliation motivation scores and academic achievement.

In Table 4.7, the correlations of the motivation scales and academic achievement are presented.

---

[1]Mitchell, op. cit.

[2]Broverman, et al., op. cit.

[3]Weiss, et al., op. cit.

[4]McClelland, op. cit.

TABLE 4.7

CORRELATIONS AND SIGNIFICANCE OF CORRELATIONS OF MOTIVATION SCALES AND ACADEMIC ACHIEVEMENT FOR 54 HIGH SCHOOL STUDENTS

	1963 GPA			1964 GPA		
Method	N-Achievement	N-Affiliation	t	N-Achievement	N-Affiliation	t
TAT	-.067	-.060	-.033	-.098	-.024	.349
TOI	.207	-.139	1.67	.200	.032	.806
EPPS	.227	.012	1.06	.360	-.117	2.49*
Self-Report	.279	-.101	2.85*	.301	-.149	3.48*

*t value significant at .05 level

Hypothesis 5 is given considerable support by these findings. In all cases, the n-achievement variable is more highly related to academic achievement than is the n-affiliation variable, although in the case of the TAT, the relationship is negative. However, the differences between the coefficients reached statistical significance only in three cases (the Self-Report and the EPPS n-achievement scales).

The correlation between the n-affiliation scales and achievement while very low was nearly consistently negative. Six of the eight obtained coefficients were negative. The significance of this finding is limited by the size of the relationship, however, it is consistent with the relationship of n-affiliation to n-achievement and gives further evidence of its possibly negative relationships with n-achievement.

Hypothesis 6 states that the relationship of achievement motivation scores to intelligence test scores will be lower than the relationship of achievement motivation scores to academic achievement. This hypothesis is derived from the ordinary definitions which are applied to the term intelligence. This hypothesis makes the probably unwarranted assumption that intelligence tests fulfill some of these definitions.

The hypothesis rests on the belief that academic achievement behavior is more a function of the energizing

and driving aspects of motivation than is intelligence test performance. Admittedly, both involve competition with a standard of excellence and possibly unique accomplishment, but the other elements of achievement motivation do not intrude as heavily into the intelligence test performance as they do into the longer behavior of the classroom. The conceptual level is of course far removed from the realities of the classroom. Further assumptions are needed about the relevance of rewards (grades) to level of achievement behavior, and the consistency of intelligence tests to their definitions. The conception of intelligence as ability and potentiality for abstract problem solving does not preclude motivational factors. The intense activity of the test taker also requires motivation for its completion. This hypothesis merely asserts that the motivational relationship to academic achievement behavior will be greater than to intelligence test performance.

TABLE 4.8

COMPARISONS OF CORRELATIONS OF ACHIEVEMENT MOTIVATION AND INTELLIGENCE AND ACHIEVEMENT SCORES

Motivation Scale	1964 GPA N = 54	Intelligence N = 50	"t" Value
TOI	.200	-.217	3.15*
TAT	-.098	.089	-1.31
Self-Report	.301	.167	0.97
EPPS	.360	.286	0.55
M-Scales	.356	.388	- .24

*t value significant at .05 level

The data bearing on this hypothesis is presented in Table 4.8. The hypothesis is not strongly supported by the evidence presented here. Only one of the five "t" ratios of correlation differences is statistically significant, and more importantly, two of the five -ach tests are more postively related to intelligence (the TOI and the M-Scales).

The above finding presents certain interpretive difficulties. The theoretical expectation made about motivation and these variables clearly does not hold.

Two factors may have produced the above finding. The Lorge-Thorndike may be more highly related than other intelligence tests to motivational factors. On the other hand, academic achievement behavior may not be as highly related to achievement motivation as previously believed. The recent formulations of Atkinson[1] would indicate that a straight-forward relationship between n-ach and achievement behavior is not always expected.

These findings confirm previous research by Krumboltz and Farquhar[2] and Farquhar[3] who found the achievement motivation scores to be as highly related to intelligence scores as to achievement scores.

---

[1]Atkinson, op. cit., An Introduction . . . .

[2]Krumboltz and Farquhar, op. cit.

[3]Farquhar, op. cit.

Hypothesis 7 states that affiliation motivation scores will be more highly related to academic performance than they will be related to intelligence.

The expectation was that need to affiliate would be expressed more openly, and visably in the classroom and would thus be more highly related to teachers' perceptions of achievement (whether negatively or positively) than it would be related to intelligence test performance. Table 4.9 contains the correlations of affiliation motivation scores and intelligence and achievement.

TABLE 4.9

CORRELATIONS BETWEEN AFFILIATION MOTIVATION SCORES AND GRADE POINT AVERAGE AND LORGE-THORNDIKE SCORES FOR FOUR AFFILIATION MOTIVATION TESTS

N-Affiliation Method	GPA N = 54	Lorge-Thorndike N = 50
TOI	.032	-.011
TAT	-.024	-.335*
Self-Report	-.148	-.300*
EPPS	-.118	.005

Probability r = 0 $<$ .05

Contrary to expectations, the relationship of affiliation motivation to intelligence test performance is higher (although more negative) than its relationship to academic achievement.

One possible explanation of this unanticipated finding is that persons with high needs to affiliate attend to stimuli which are unimportant for high I.Q. test performance. That is, they spend their time expressing their needs for affiliation rather than engaging in achievement behavior. That is, a greater amount of time is spent in non-intellectual behavior and this leads to poorer I.Q. test performance.

The reason for the more highly negative relationship to intelligence cannot be explained on the same basis. Possibly affiliative behavior is partially rewarded by the teacher, which counteracts the effects of the lower time spent on learning tasks (which is also likely to be rewarded in school).

There is some evidence that anxiety and needs for affiliation are positively related.[1] The further finding by Wrightsman[2] that anxiety and I.Q. are negatively related ($r = -.37$), would be consistent with the above findings.

---

[1]C.N. Cofer and M.H. Appley, Motivation: Theory and Research, John Wiley and Sons, Inc., New York, 1964. See chapter 14 for a discussion of anxiety and affiliation.

[2]Lawrence S. Wrightsman, Jr., "The Effects of Anxiety, Achievement Motivation and Task Importance Upon Performance on an Intelligence Test," Journal of Educational Psychology, Vol. 53, No. 3, 150-156.

Factor Analysis of Test Scores

A principal axis factor analysis of the nine test intercorrelation matrix was performed. The hypothesis analyzed was that the n-achievement tests would all have high loadings on one factor and the n-achievement tests on a second factor. The unrotated factor loading matrix is reproduced in the appendix. Two factors were rotated by both the varimax and quartimax methods, and produced similar results, which are shown in Table 4.10.

TABLE 4.10

ROTATED FACTOR LOADINGS OF FIVE ACHIEVEMENT MOTIVATION AND FOUR AFFILIATION MOTIVATION SCALES N = 112

Test Variable	Factor 1	2
M-Scales	.758	.035
TOI N-achievement	.374	.050
TOI N-affiliation	.024	.122
TAT N-achievement	.304	-.095
TAT N-affiliation	-.081	.363
Self-Report N-achievement	.760	.014
Self-Report N-affiliation	.084	.591
EPPS N-achievement	.483	-.514
EPPS N-affiliation	-.162	.474

The factor analysis confirms the convergent and discriminant validity hypotheses formulated and examined earlier (Hypotheses 1 - 3). The tests have loadings according to expectations although the n-affiliation loadings are quite small. However, most of the relationships of n-affiliation throughout the study were very low.

CHAPTER V

SUMMARY AND CONCLUSIONS

Major Findings

The main concern of the study was the analysis of the construct validity of five achievement motivation tests. The interrelationships of these tests of achievement motivation, their relationship to measures of affiliation motivation, and to other behavior was the method employed to analyze the construct validity of these tests. While the main focus of the study was on achievement motivation, the procedures used also yielded information about affiliation motivation, academic achievement and intelligence. The main findings are listed.

1. The five achievement motivation scales are moderately interrelated ($r = .156$ -- $r = .597$), and thus convergent validity of these scales was demonstrated.
2. The projective achievement motivation (TAT, TOI) scales are generally correlated with other variables to a lesser degree than the more objective measures (EPPS, M-Scales).
3. The affiliation motivation scales are, in general, unrelated to each other ($r = -.030$ -- $r = .247$),

and the convergent validity requirement is not met to any degree.

4. The achievement motivation scales are more positively related to each other than to measures of affiliation motivation. The relationships of the two traits is in general negative, although this relationship is partly spurious due to methods factors in the tests.
5. A short one-item self-rating achievement motivation measure was more highly related to academic achievement score than were the projective measures of achievement motivation. The self-rating device was only slightly inferior in prediction to the EPPS and the M-Scales.
6. The test-retest reliability of the total score of the M-Scales was r = .82. The test-retest period was ten weeks. The subtest reliability was lower for all scales except the Word Rating List (r = .30 -- r = .70) the reliability of the Word Rating List was r = .86.
7. Affiliation motivation measures were not substantially related to academic achievement (r = .089, r = -.149). Most of the scales were negatively related to achievement.
8. Achievement motivation scales were all positively related to academic achievement except for the TAT

which had an insignificant negative correlation with academic achievement. These positive correlations ranged from r = .20 to r = .375, with the EPPS and the M-Scales being the most highly related.

9. Achievement motivation scales are also moderately related to intelligence test scores, except for the TOI which was insignificantly negatively related to intelligence test scores. The correlations ranged from r = -.216 to r = .389.

10. Affiliation motivation scores tended to be negatively related to intelligence test performance. Two of these relationships were statistically significant (TAT and Self-Report, r = -.335, and r = -.298).

11. A principal axis factor analysis of the test score matrix with a varimax and quartimax rotation of two factors produced a set of factor loadings in which the n-achievement scales loaded most heavily on the second factor. The Self-Report n-achievement scale and the M-Scales had the highest rotated factor loadings on the first factor, while the Self-Report n-affiliation scale and the EPPS n-affiliation scale weighted most heavily on the second factor in both rotational procedures.

Discussion of the Major Findings

At the beginning of the study, the main purpose was conceived to be an analysis of the construct validity of the Michigan State M-Scales. However, as the problem became more clearly enunciated, it was evident that not only was the M-Scales on trial, but all other motivational instruments were as well. This being the case, the validity problem for all instruments must be considered more or less spontaneously. As Campbell[1] stated, validation in this case is "symmetrical and equalitarian," and all tests receive confirmatory and disconfirmatory validity status at the same time.

There are, however, analytic procedures in this study which seem to isolate the instruments which are not clearly living up to their billing. If nearly all the relationships of a certain variable with others are low and contrary to prediction, there is clearly something wrong with the instrument and its application. There was also the relationship of the motivational tests to motive-relevant behavior. Here, also, a lack of predictable relationships would tend to incriminate one of the instruments, especially if other tests fulfill the theoretical predictions.

The main purpose of the study became then to examine the construct validity (trait variety) of five types of n-achievement scales, and four n-affiliation scales.

---

[1]Campbell, op. cit.

Secondarily and concomitantly, the construct of achievement motivation was also being investigated. Clearly, these two analyses occur concurrently. Had none, or few of the relationships fulfilled the theoretical expectation, both the tests and the construct would require further examination. The case was, however, one of mixed supporting and informing evidence.

The summary will first of all consider the validating and invalidating information produced by the investigation; secondly, it will consider experimental procedures possibly responsible for these findings, and will conclude with implications of the study and recommendations about further research.

The method of this study was entirely correlational. The size and sign of the correlation coefficient was the primary source of information in the study.

Validating and Invalidating Evidence

The evidence for the construct validity of the n-achievement scales is mainly contained within the multitrait-multimethod matrix. Except for a few cases in which the predicted ranking was reversed, the matrix tended to be structured the way it should if all the tests of n-achievement were measuring the same underlying construct. That is, it was demonstrated that the convergent validity requirement was met to some degree.

Within the matrix, at least, the divergent validity requirement was also met to some degree. The intercorrelations

of n-achievement scales were usually higher than the correlations of n-achievement scales with n-affiliation scales.

Furthermore, the correlations of n-achievement scales and academic achievement (achievement oriented behavior) was in general positive and statistically significant. These correlations with academic achievement were also higher than the correlations of n-affiliation and academic achievement.

A factor analysis of the test score matrix produced a set of factor loadings which was generally consistent with the predicted result. However, the loadings were moderate, especially for the projective type instruments.

There was also much contradictory evidence. Most crucial were the extremely low correlations produced within the study. Although the multitrait-multimethod matrix was generally structured in the predicted direction, the obtained correlations were so low as to be practically insignificant. While this was not a hoped-for result, it was consistent with other studies which investigated these relationships with other samples, and found disappointingly low intercorrelations.

There was also the relationship of the one item self-report inventory to academic achievement and intelligence. In most cases the correlations obtained with this measure were as high as those with the more complex measures. While this is perhaps evidence more for the lack of utility of the

more complex instruments, it is also evidence that the traits being measured may be more accurately measured. The one-item test would be expected to be somewhat unreliable and thus relate to other variables in an attenuated fashion.

There was also the high correlations obtained between the n-achievement scores and the intelligence scores. Although one n-achievement measure was more significantly related to GPA than to I.Q., others did not have this differentiation.

The factor analysis was also a mixed blessing. The loadings of the projective n-achievement measures on the first factor were very low, and the most widely used n-achievement measure (TAT) had its highest positive loading on a third factor. The second factor was positively loaded with the n-affiliation tests.

The evidence of construct validity of these tests was therefore both confirming and disconfirming. The evidence considered in total would indicate that there is possibly something being measured in common by all n-achievement scales, but that at least for the projective measures of n-achievement, there is also something very unique being measured. This conclusion must be tempered, however, in light of the procedures used in the study. These are presented and discussed in the following section.

Experimental Procedures Possibly Distorting the Results of the Study

The choice of a high school sample may have influenced the results. Although some studies have used the TAT method on high school samples, the M-Scales was the only test which has been specifically validated on high school populations. The high school sample, however, did consistently produce scorable stories. The mean and standard deviation of n-achievement scores, moreover, were consistently higher than those reported by Atkinson, et al.,[1] for older samples. However, the n-affiliation scores obtained from the TAT scoring were much lower than those obtained in other studies. The means and standard deviations of the TOI n-achievement and n-affiliation scores were not reported in Atkinson, et al.,[2] but the means and S.D.'s of the present study were likely to be lower. The selection of high school samples may then have seriously affected the elicitation of affiliation imagery. Reasons for this could lie in the nature of the testing session. The tests were administered within the school setting where extra-test cues would be higher for achievement than for affiliation (this is, however, a debatable point).

The mixing of the eleventh and twelfth grade students in sample C, may also have influenced the relative prediction of GPA. The fact that both the M-Scales and EPPS n-achievement were moderately related to the GPA would

---

[1]Atkinson, et al., op. cit., "The Achievement . . .

[2]Ibid.

obviate somewhat the conclusion that the lack of prediction of the projective n-achievement scales was due entirely to this heterogeneity. In any case, high school samples would have sufficient variablity in ability so that all relationships would not be severely attenuated. It is also obvious that the subjects taken by eleventh and twelfth grade students differ. However, the rather high intercorrelation between eleventh and twelfth grade GPA ($r = .84$) would indicate that this is not a completely invalidating condition.

There is also the post-dictive nature of the relationships of motivational scales and intelligence and academic achievement. Both the GPA and intelligence scores are based on behavior samples obtained prior to the motivational testing. While this likely leads to the lower correlations consistently exhibited in the study, all these traits are considered to be sufficiently stable in time so that post-dictions would be similar to true predictions.

The ipsative nature of the EPPS and the possible ipsative result of the projective scoring may also be a significant factor. The ipsative quality of the EPPS results from the forced-choice method of testing. The negative correlation of n-achievement and n-affiliation results from method of the projectives. For any given fixed amount of content elicited, the scoring of this content for

n-achievement probably limits the amount of scorable imagery for n-affiliation. The knowledge of the net impact of these somewhat artificial negative correlations would await a study in which these factors are controlled.

The variations in testing conditions throughout the three school systems undoubtedly influenced the results, especially for the projective tests. Examination of the correlation matrices for the three samples shows definite variation in the patterns of intercorrelation. In one sample the TAT-motivational scores are much more highly related to other variables than for the total group. While sampling variation of correlation coefficients is highly likely with small samples, the amount of reversals in the sub-samples is likely due to some of the experimental arrangements. While no justification for these arrangements is offered, this condition has to be acknowledged as a contaminating factor in the study.

The basic psychometric qualities of the TAT and TOI have to be considered. Atkinson has marshalled evidence that the serial position of the pictures affects the predictive validity of these tests. These serial position effects are believed to be due to a "set of response variability"[1] which seriously limits the imagery available after four pictures. This same phenomenon would likely be in evidence when two projective type instruments are administered

---

[1] Ibid.

in succession. Other studies have not controlled this factor and may also have attenuated relationships between n-achievement scores and other variables.

Finally, there was the variation in the pictures used to elicit thematic imagery. While the pictures were the same, the clarity was less than the original pictures. The consistent elicitation of achievement imagery would indicate that the pictures are perceptually interpreted, there is the possibility that this was a factor in the lack of production of affiliation imagery. However, Atkinson has demonstrated that predictive validity coefficients obtained by pictures judged relevant to achievement fantasy elicitation were no better than pictures judged relevant to affiliation imagery when predictions were made about achievement oriented behavior. Similar reasoning would indicate that as long as imagery is elicited, the clarity and content of the picture is not a very important factor.

Recommendations and Implications of the Study

The results of the study and the discussion of the possible meaning of these results lead to a set of recommendations about needed research and possible additional requirements of theories of achievement and affiliation motivation.

1. In the first place, the low interrelationships obtained within this study should be cross-validated

on other more carefully chosen samples, before any assertion is made that these tests are measuring similar traits.

2. The length of projective tests should be varied to see whether increased items can elevate the stability reliability of these kinds of instruments. Although four items have been shown to be an optimum for predictive validity requirements, this number of items may simply be too few to reliably measure the trait.
3. The tendency for investigators to use group differences as a criterion for validity of projective *n*-achievement scales may seriously mask the low underlying relationships obtained under correlational methods. While group differences are adequate beginnings for test validation, the conceptions presently advanced about the -achievement trait indicate that adequate correlations should ultimately be obtained when using -achievement as an independent variable.
4. The multitrait-multimethod matrix, as a method of test validation should include statistical tests of correlations as an overall evaluation of trait validity.
5. When interrelationships of traits are studied, artificial relationships, such as those obtained by the EPPS, should be removed as much as possible.

6. The possible artificial negative correlations produced by multiple scorings of projective instruments should be reduced as much as possible. An alternative procedure such as separate sets of pictures, used with adequate test intervals, scored only for n-achievement and n-affiliation, would eliminate some of the spurious negative correlation.
7. Large data samples should be used for this kind of work. The labor and cost of scoring of sample sizes needed for proper analysis, are too prohibitive for separate studies outside the larger programmatic studies.

There are also a number of implications for theory development which emerge from the study.

8. Broverman's[1] conception of fantasy as an alternative to the behavioral expression of achievement motivation should be seriously examined as an explanation of the relationship of fantasy to achievement oriented behavior. The results of the study indicate that TAT n-achievement scores are only moderately related to one kind of achievement behavior, and there is the slight suggestion that this may be a negative relationship.

---

[1]Broverman, Jordan and Phillips, op. cit.

9. The insertion of a n-achievement term in theoretical equations may await a more reliable measurement of the trait represented by this term. While there is no finality in the statements made about n-achievement scores produced by scoring schemes such as those in Atkinson,[1] there is the implicit suggestion that these scoring schemes produce stable scores as well as scores which have high inter-rater consistency.
10. Conceptions of intelligence may have to include motivational components. The rather consistent relationships of n-achievement scores to the I.Q. measure would indicate that intelligence test performance may be more highly weighted with motivational components, than was previously believed.
11. Affiliation motivation may have to be considered in newer formulations of so-called achievement oriented behavior. The somewhat consistent negative relationship may only be a result of the experimental arrangements, but it may also add predictable components to any new conceptions of achievement oriented behavior.
12. The unitary trait conception of n-achievement may have to be modified. While the evidence is certainly

---

[1]Atkinson, et al., op. cit., "The Achievement . . .

not compelling, there are hints that n-achievement may be multi-faceted, some facets relating to achievement behavior, others being unrelated.

In conclusion, two statements from Atkinson are highly informative as to the general condition of the construct of achievement motivation. In 1958 he and Reitman wrote:

> "It is clear however that studies using thematic apperceptive measures can contribute little to real psychological progress until such time as these studies proceed from a firm factual knowledge of the basic intrinsic properties of the instruments." [1]

In 1964 he stated:

> "Research on achievement motivation has come a long way from the earliest studies which proceeded to explore the simple hypothesis that TAT n-achievement scores should be positively related to achievement oriented behavior."[2]

These statements clearly illustrate the dilemma of investigators who wish to use the projectively measured n-achievement construct in their studies of behavior. On the one hand, theoretical analyses have progressed very much, but the "real psychological" progress may have to await "a firm factual knowledge of the basic intrinsic properties of the instruments." It is obvious that one cannot proceed without the other.

---

[1] Ibid.

[2] Atkinson, op. cit., An Introduction . . . .

The more objective instruments clearly relate more highly in predicted directions than do the more elaborate projective tests. However, even these are not much more highly successful than a short self-rating measure of the n-achievement trait.

BIBLIOGRAPHY

American Psychological Association, Committee on Psychological Tests. Technical Recommendations for Psychological Tests and Diagnostic Techniques. Washington, D.C.: APA, 1954.

Atkinson, John W. An Introduction to Motivation. D. Van Nostrand and Co., Inc., 1964.

Atkinson, John W. (ed.) Motives in Fantasy Action and Society. D. Van Nostrand Co., Inc., 1958.

Atkinson, John W., et al. "The Achievement Motive, Goal Setting and Probability Preferences," Journal of Abnormal and Social Psychology, 60, No. 1, 1960, 27-36.

Bechtoldt, Harold. "Construct Validity: A Critique," American Psychologist, Vol. 14, No. 10, October, 1959, 619-629.

Bendig, A.W. "A Preliminary Factor Analytic Investigation of Need Achievement Items," Journal of Psychological Studies, 11, 1959, 32-38.

Bendig, A.W. "Manifest Anxiety and Projective and Objective Measures of Need Achievement," Journal of Consulting Psychology, 21, 1957, 354.

Bernardin, A.C. and Jessor, R. "A Construct Validation of the Edward's Personal Preference Schedule with Respect to Dependency," Journal of Consulting Psychology, 21, 1957, 63-67.

Birney, Robert C. "The Reliability of the Achievement Motive," Journal of Abnormal Social Psychology, 58, 1959, 266-267.

Blommers, Paul and E.F. Lindquist. Elementary Statistical Methods. Boston: Houghton Mifflin Co., 1960, 451.

Broverman, Donald M., Jordan, Edward J., and Phillips, Leslie. "Achievement Motivation in Fantasy and Behavior," Journal of Abnormal and Social Psychology, 60, No. 3, 1960, 374-378.

Burdick, Harry. "Interrelationships of Three Measures of Motivation," Psychological Reports, 8, 1961, 225-226.

Campbell, D.T. "Recommendations for APA Test Standards Regarding Construct Trait or Discriminant Validity," American Psychologist, Vol. 15, No. 8, August, 1960, 546-553.

Campbell, D.T. and Fiske, D.W. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix," Psychological Bulletin, Vol. 56, No. 2, 1959, 81-105.

Cole, D., et al. "The Relation of Achievement Imagery Scores to Academic Performance," Journal of Abnormal and Social Psychology, Vol. 65, No. 3, 1962, 208-211.

Cronbach, Lee J. and Meehl, Paul E. "Construct Validity in Psychological Tests," Psychological Bulletin, Vol. 52, No. 4, July, 1955, 281-302.

Cureton, Edward. "Validity," in E.F. Lindquist, Educational Measurement. Washington, D.C.: ACE, 1951, 621-692.

Dicken, Charles F. "Convergent and Discriminant Validity of the California Psychological Inventory," Educational and Psychological Measurement, Vol. 23, No. 3, 1963, 449-459.

Ebel, Robert L. "Must All Tests Be Valid?," The American Psychologist, 1961, 640-647.

Edwards, A.L. Personal Preference Schedule Manual. The Psychological Corporation, 1954.

Farquhar, W.W. Motivation Factors Related to Academic Achievement. United State Office of Education Cooperative Research Project No. 846, 1963.

French, Elizabeth G. "Some Characteristics of Achievement Motivation," Journal of Experimental Psychology, Vol. 50, No. 4, 1955, 232-236.

French, Elizabeth G. and Chadwick, Irene. "Some Characteristics of Affiliation Motivation," Journal of Abnormal and Social Psychology, Vol. 52, No. 3, 1956, 296-300.

Green, Robert Lee. "The Predictive Efficiency and Factored Dimensions of the Michigan State M-Scales for Eleventh Grade Negro Students, An Exploratory Study," Thesis for the Degree of Ph.D., Michigan State University, 1962.

Hempel, C.G. "Fundamentals of Concept Formation in Empirical Science," International Encyclopedia of Unified Science, Vol. II, No. 7, Chicago: University of Chicago Press, 1952.

Herron, E. Wayne. "Intellectual Achievement Motivation: A Study in Construct Clarification," The University of Wisconsin-Milwaukee, 1964.

Hills, John R. "Needs for Achievement, Aspirations, and College Criteria," Journal of Educational Psychology, Vol. 41, No. 1, 1958, 156-161.

Himelstein, R. and Kimbrough, W.W., Jr. "Reliability of French's 'Test of Insight'," Educational and Psychological Measurement, Vol. 20, No. 60, 1960, 737-741.

Himelstein, P., Eschenbach, A.E., Carp, A. "Interrelationships Among Three Measures of Need Achievement," Journal of Consulting Psychology, Vol. 22, No. 6, 1958, 451-452.

Jessor, R. and Hammond, K.R. "Construct Validity and the Taylor Anxiety Scale," Psychological Bulletin, Vol. 54, 1957, 161-170.

Johnson, Van C. "An Assessment of the Motivation Factor in the Estimation of Academic Achievement of Eleventh Grade Indian Students and the Factored Dimensions of the M-Scales." An Exploratory Study, An unpublished doctoral dissertation, Michigan State University, 1963.

Jordan, T.E. and DeCharms, R. "The Achievement Motive in Normal and Mentally Retarded Children," American Journal of Mental Deficiency, 64, 1959, 457-466.

Kagan, Jerome and Moss, Howard A. "Stability and Validity of Achievement Fantasy," Journal of Abnormal and Social Psychology, Vol. 58, No. 3, May, 1959.

Karolchuk, P. and Worell, L. "Achievement Motivation and Learning," Journal of Abnormal and Social Psychology, 53, 1956, 255-257.

Kausler, D.H. and Trapp, E.P. "Methodological Considerations in the Construct Validation of Drive Oriented Scales," Psychological Bulletin, 56, 1959, 152-157.

Kipfmueller, Mark K. "The Predictability and Factored Dimensions of the M-Scales for Eleventh Grade Parochial School Students," unpublished doctoral dissertation, Michigan State University, 1963.

Krumboltz, J.D. and Farquhar, W.W. "Reliability and Validity of the N-Ach Test," Journal of Consulting Psychology, Vol. 21, No. 3, 1957, 226-228.

Littig, Lawrence W. and Yeracaris, Constantine A. "Academic Achievement Correlates of Achievement and Affiliation Motivations," Journal of Psychology, 55, No. 1, 1963, 115-119.

Loevinger, Jane. "Psychological Tests as Instruments of Psychological Theory," Psychological Reports, 1957, 635-694.

Longnecker, E.D. "Perceptual Recognition As a Function of Anxiety Motivation, and the Testing Situation," Journal of Abnormal and Social Psychology, 64, 1961, 215-221.

Marlowe, D. "Relationships Among Direct and Indirect Measures of the Achievement Motive and Overt Behavior," Journal of Consulting Psychology, Vol. 23, No. 4, 1959, 329-332.

McClelland, D., Atkinson, J., et al. The Achievement Motive. New York: Appleton-Century-Crofts, 1953.

Melikian, Levon H. "The Relationships Between Edward's and McClelland's Measures of Achievement Motivation," Journal of Consulting Psychology, Vol. 22, No. 4, 1958, 296-299.

Mitchell, James J. "An Analysis of the Factorial Dimensions of the Achievement Motivation Construct P," Journal of Educational Psychology, 1961, 179-187.

Murray, H.A. Exploration in Personality. New York: Oxford Press, 1938.

Payne, David A. "The Concurrent and Predictive Validity of an Objective Measure of Academic Self-Concept," Educational and Psychological Measurement, 1962.

Phares, E.J. and Adams, C. "The Construct Validity of the Edwards PPS Heterosexuality Scale," Journal of Consulting Psychology, 1961, 25, 341-344.

Reitman, Walter R. "Need Achievement, Fear of Failure, and Selective Recall," Journal of Abnormal and Social Psychology, Vol. 62, No. 1, 1961, 142-144.

Riccuiti, Henry N. "The Prediction of Academic Grades with a Projective Measure of Achievement Motivation I Initial Validating Studies," Princeton, New Jersey, ETS, 1954.

Shaw, M.C. "Need Achievement Scales as Predictors of Academic Success," Journal of Educational Psychology, Vol. 52, No. 6, 1961, 282-285.

Thorpe, Marion D. "The Factored Dimensions of an Objective Inventory of Academic Motivation Based on Eleventh Grade Male Over- and Under-achievers," unpublished doctoral dissertation, Michigan State University, 1961.

Uhlinger, C.A. and Stephens, M.W. "Relation of Achievement Motivation to Academic Achievement in Students of Superior Ability," Journal of Educational Psychology, Vol. 51, No. 5, 259-266.

VanZandt, B.R., and Himelstein, P. "The Role of Verbal Fluency on a Projective Measure of Motivation," Educational and Psychological Measurement, Vol. 21, No. 4, 1961, 873-878.

Walker, H.M. and Lev, J. Statistical Inference, Henry Holt and Co., New York, 1953.

Weiss, P., Wertheimer, M. and Groesbeck, Byron. "Achievement Motivation, Academic Aptitudes, and College Grades," Educational and Psychological Measurement, Vol. 19, No. 4, 1959, 663-666.

Zuckerman, M., et al. "Concurrent and Construct Validity of Direct and Indirect Measuring Dependency," Journal of Consulting Psychology, Vol. 25, No. 4, 1961, 316-323.

APPENDIX A

THE MATRIX OF FACTOR LOADINGS PRODUCED BY A PRINCIPAL AXIS ANALYSIS OF FIVE N-ACHIEVEMENT AND FOUR N-AFFILIATION TEST SCORES FOR 112 STUDENTS

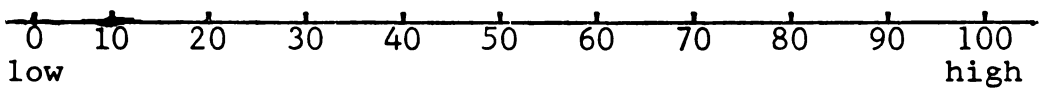
		Factor Number								
		1	2	3	4	5	6	7	8	9
Variable Number	1	.711	.009	.021	-.125	.264	-.180	-.108	-.000	.159
	2	.324	.180	.095	.360	.192	.023	-.062	.103	-.033
	3	-.025	-.041	.213	-.251	.121	.047	-.018	.115	-.103
	4	.317	-.121	.082	.128	.031	.027	.460	-.200	.089
	5	-.217	-.159	.036	.131	.301	.048	-.432	-.177	.066
	6	.711	.063	-.079	-.140	.287	.122	-.028	-.158	-.159
	7	-.154	-.137	-.071	.078	.576	-.044	.176	.201	-.128
	8	.646	-.145	-.032	.043	-.283	.105	-.094	.250	.103
	9	-.335	.105	-.022	-.111	.372	.114	.104	.071	.276

APPENDIX B

SELF REPORT INVENTORY

After reading the statement below, rate yourself on the scale as to how well you fit the description in the statement. If you feel that the statement characterizes you very well, you should mark at the high end; if you feel the statement is not typical of you, then you should mark at the lower end. If you are somewhat similar, you should mark in the middle portion.

I like to do my best, to be successful, to accomplish tasks requiring skill and effort, to be a recognized authority, to accomplish something of great significance, to do a difficult job well, to solve difficult problems and puzzles, to be able to do things better than others, to write a great novel or play.



I like to be loyal to to friends, to participate in friendly groups, to do things for friends, to form new friendships, to make as many friends as possible, to share things with friends, to do things with friends rather than alone, to form strong attachements, to write letters to friends.