THE APPLICATION OF A RATIO SCALING TECHNIQUE TO MEASUREMENT OF THE ACHIEVEMENT MOTIVE

Thesis for the Degree of M. A.
MICHIGAN STATE UNIVERSITY

David P. Goff

1969

ABSTRACT

THE APPLICATION OF A RATIO SCALING TECHNIQUE TO MEASUREMENT OF THE ACHIEVEMENT MOTIVE

By

David P. Goff

The lack of mathematically precise quantification of personality variables has always imposed limitations on the "scientific" status of the personality-clinical fields of psychology. The projective techniques represent, perhaps, one of the most extreme deviations from objectivity in measurement. The purpose of this study was to select a unidimensional ratio scaling model suitable for employing TAT protocols as the stimulus material and test it by measuring a simple personality attribute, the need for achievement.

The data (protocol sets) to be scaled were selected from a successful cross-validation study (Klinger, 1967) of McClelland's original achievement arousing experiments. TAT protocol sets from four subjects were sampled from each of two experimental groups, "neutral" and "achievement aroused," which differed significantly on the basis of the n Ach scores obtained via McClelland's scoring system. The outcome of Klinger's study was to be used as the external validity criterion for assessing measurement precision of the scaling model.

The model selected was a variant of a direct-estimate fractionation method developed by Comrey (1950) and subsequently modified by O'Grady and

Karon (1968) and Carter (1966). A judge's task, using an abridged version of McClelland's scoring criteria, was to decide which of two protocol sets reflected the most need achievement and assign to it the entire length of a 20 cm. line. The relative magnitude of the smaller member of the pair was indicated by placing a mark some place along the line. The ratio of the two distances was obtained for all possible paired-comparisons of the eight stimuli (subjects). From the resulting matrix of direct ratio estimates, scale values were obtained for each of the stimuli. The scaling procedure was repeated independently from the ratings of four separate judges.

Because the method here described differs from the more classical procedures in two ways, i.e. the method of obtaining the direct ratio estimates and the use of a single judge to obtain the raw data, an independent justification of the computational procedures was required. Consequently, a least-squares solution for the scale values and a test for internal consistency were derived for the present case. The latter index is a measure of the degree to which the observed data are internally consistent with the theoretical assumptions of the model.

Four questions were posed by the study. Does the TAT measure of the achievement motive prove scalable, i.e. do the data fit the model? Will the judges agree with each other? Will the scale values of the judges measure the same attribute as the original McClelland scores? Does the scaling technique provide an increase in measurement precision compared with the McClelland system?

The answers to the first three questions were affirmative. The average internal consistency coefficient was 0.94. With the exception of one judge (the most "naive"), the inter-judge reliabilities were high,

0.88 to 0.95, and the concurrent validity coefficients were from 0.86 to 0.91. However, an increase in measurement precision was not demonstrated, i.e. the degree of separation between the treatment groups afforded by the new scale values was not greater compared with the McClelland scores.

A modification of the study was proposed, whereby each of the four TAT stories in a protocol set would be scaled separately. It was hypothesized that an arithmetic average of scale values across stories would result in more accurate prediction of the criterion groups. Procedures for additional verification of the ratio properties of the scale values were also suggested.

Approved: J. S. Helmann

Date 5/16/69

THE APPLICATION OF A RATIO SCALING TECHNIQUE TO MEASUREMENT OF THE ACHIEVEMENT MOTIVE

By

David P. Goff

A THESIS

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

MASTER OF ARTS

Department of Psychology

1969

DEDICATION

To Anna-Karin

ACKNOWLEDGMENTS

I am most grateful to all of the people, including those not connected with my project, who contributed to my understanding of the relevant issues of this study. I am particularly indebted to Dr. James Uleman, committee chairman, whose patience, support and permissive open-mindedness gave free vent to the development and exploration of my own interests; factors which contributed immeasurably to my eagerness and spirit of total involvement. My warm thanks are given to Dr. Charles Hanley, committee member, whose insightful criticisms resulted in the final formulation of the experimental design. I am also grateful to Dr. Terrence Allen, committee member, Dr. Bertram Karon, whose previous work formed the backbone of this project, and Dr. Jack Hunter for their many hours of valuable consultation time.

Finally, I thank my wife, Anna-Karin, who gave unfailing support and who poured heart and soul into the difficult typing of this manuscript.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

An important criticism frequently directed to the fields of personality and clinical psychology is the failure to employ more sophisticated and mathematically precise techniques to quantify personality variables. The projective techniques which, of necessity, compromise objectivity for maximum freedom of subject response, are perhaps the best cases in point. A significant problem area has been the scale of measurement attainable and, of course, the host of scientific restrictions resulting from this limitation. Ordinal scale measurement comprises, in most cases, the highest precision level attained and is well exemplified by the line of research initiated by McClelland, Clark, Roby, and Atkinson (1949) with the TAT.

It was the purpose of this study to select and test a unidimensional ratio scaling model which might prove workable with projective materials--particularly the TAT. Most desirable was the choice of a model which maintains, to the greatest extent possible, the psychometric rigor of the psychophysical models, while at the same time permitting judgment operations that would be feasible with TAT type stimuli.

Selection of a Model

It must be noted at the outset that this study is not concerned with modifications of already existing methods for administering stimulus cards or obtaining TAT protocols, i.e. any alterations in the free

response features of the test situation itself. Rather, the starting point will be with the protocols. A set of protocols for a particular subject will, henceforth, be referred to as the "stimulus" which is to be assigned a numerical scale value to represent the amount or magnitude of some unidimensional attribute presumably contained by the subject.

The complexity of these stimuli and of even the most simple personality variables (e.g. psychologial needs) seems to place rather strict limitations on the choice of a suitable model from among those that customarily deal with simpler stimuli or variables. One such limitation is that the number of available judges having both the academic background and the available training time to rate material of this very specific nature is relatively small. Thus, a "quantitative-judgment" method requiring judgments of relationships among psychological distances or ratios, was preferred to a "variability-judgment" method which requires a large number of ordinal type judgments for each comparison.[1] Needless to say, the latter is a plausible alternative that could utilize the combination: large number of relatively naive judges and simpler (ordinal) type ratings. Subjective anchoring is another difficulty which increases with the complexity of the stimuli. For this reason, the "subjective-estimate" methods, requiring a direct magnitude estimation with respect to a constant origin, were rejected in favor of the "fractionation-methods" which require the less complicated task of using one stimulus as a standard for comparison with another. A third limitation to be dealt with is the fact that the stimuli as defined above, are

---

[1]The terminology used here refers to Torgerson's scaling classification scheme. It is only one of many but seemed the most relevant for present purposes. The interested reader is referred to Torgerson (1958), Chapters 3 and 5.

neither adjustable nor do they have a related attribute as a basis for pre-determining prescribed ratios--like, for example, intensity of light or frequency of sound. Since these conditions are precisely stipulated by the "prescribed-ratio" methods, the alternative "direct-estimate" approaches are clearly indicated. In the latter, the judge is presented with two stimuli and reports the subjective ratio between them.

The model finally selected was a variant of a method developed by Comrey (1950) currently referred to as the "constant-sum method." In Comrey's method, each of N judges first divides 100 points between each of two stimuli being compared. After averaging across judges, ratios between stimuli are then obtained for all possible paired-comparisons of n stimuli where each stimulus serves as a standard for all others, giving n(n-1)/2 independently observed ratios. Scale values for the n stimuli are then computed from the ratios. As will be seen later, the logic of this model contains the inherant advantage of permitting an internal consistency test of the observed ratios.

An alternative was introduced to the above mentioned procedure for obtaining the ratios, and is based on recent studies by O'Grady and Karon (1968), and Carter (1966). These authors found that reliable ratio judgments between sets of TAT protocols could be obtained by having judges place a mark on a 20 centimeter line. Rating "emotional health" of schizophrenic patients, the judges assigned the entire length of a 20 cm. line to represent the "healthiest" member of a subject pair and then placed a mark on the line to represent the relative magnitude of the health of the sicker member. A ratio was then formed by finding the quotient of the two distances. Investigations by Estavan[1] which preceded

[1]No relevant publications.

these studies indicated that other rating procedures with TAT stimuli, for example, extending a line to indicate the ratio of the larger stimulus to the smaller, yielded unreliable ratings. A second modification introduced by the above authors was the employment of a single judge to obtain the observed ratios. Although these studies only reported concurrent type validity coefficients (0.58 to 0.71), the intra-judge reliabilities (on the order of 0.90) warranted adoption of the method for the present study. It should be noted that the mathematical features of Comrey's model are independent of the procedures for obtaining the observed ratios so long as absolute zero stands for the origin.

Experimental Design

The choice of construct or attribute, the selection of materials and the overall experimental design of this study were all based on the primary objective of maximizing the possibility of making direct inferences about measurement precision of the scaling model. The Achievement Motive, defined by McClelland, Atkinson, Clark, and Lowell (1953), was selected because of its relative simplicity as a personality variable and (hopefully) its unidimensional characteristics. In addition, definitional criteria had been empirically determined from studies involving experimenter manipulation of the independent variable (motive arousing conditions). And finally, the availability of numerous cross-validation studies of both the TAT measure and original experiment provides a realistic appraisal of error variance and a source of materials for the present study.

As part of a larger study, Klinger (1967) essentially replicated McClelland's original criterion study and obtained TAT protocols which significantly discriminated between a "Neutral" and an "Achievement aroused"

group using McClelland's scoring system "C". Copies of these protocols were gratefully obtained from Dr. Klinger and were used as the stimulus material for this study. Use of protocols from the above study now enables a direct assessment of the validity of the scaling model by holding constant such variables as success or failure of the experimental treatments, faulty administration or scoring, or failure to yield protocols which discriminate on any basis. In other words, the initial experiment has already worked and thus provides external validity criteria that are independent of the model. Considering the derived scale values as an alternate method of scoring the TAT for n Achievement, measurement precision (validity) can be assessed by comparing the new scale values with the McClelland scores in terms of the degree to which they separate the two treatment groups.

Analysis of the Model

A least-squares solution for deriving scale values from the matrix of observed ratios and a test for internal-consistency (intra-judge reliability) will be developed in the next section. After these measures are obtained for the data from this study, inter-judge reliabilities will be determined by computing the product-moment correlations between scale values of the subjects derived from the ratings of four judges. Concurrent validity will be measured by correlating the scale values with the original McClelland scores. And finally, measurement precision will be assessed by making the previously mentioned comparison between the two scoring systems in terms of the magnitude of separation between treatment groups. Specifically, the "strength of statistical association" between the independent variable, treatment conditions, and the dependent variable, n Achievement score values, will be measured for both cases.

METHOD

The Stimuli

The subjects, whose protocol sets were the stimuli to be scaled, had participated in the "Direct-Live" conditions of the Klinger study investigating "modeling effects on Achievement imagery." This phase of Klinger's experiment was a successful replication of McClelland et al.'s (1949, 1953) original Achievement-arousing studies. One group of randomly assigned subjects (n=10) received the achievement-arousing instructions developed by Haber (1957) and the other group (n=4) received a slight modification of Haber's "relaxed" instructions. The instructions were followed by a 5-minute scrambled words test and then the administration of the four-card TAT test developed by McClelland. The four stimulus cards were McClelland's original n Ach series--B, H, A, G. The four-story protocols for each subject were scored for n Ach according to the scoring manual developed by McClelland, Atkinson, Clark, and Lowell (1958). Based on the overall n Ach score for each subject, a significant difference was found between the "neutral" and "achievement-aroused" groups.

Klinger's sample sizes, though small, were ideal for the planned application of Comrey's model which requires $n(n-1)/2$ independent paired comparisons. As the number of needed comparisons increases geometrically with increasing sample size, and a total of eight stories must be read for each comparison, it can be seen that larger samples can involve a

staggering amount of work. The subjects (protocol-sets) selected as the scaling stimuli consisted of the intact neutral group (n=4) and a non-random sub-sample of n=4 from the n Ach group. The latter sub-sample was selected in accordance with the strategy of maintaining, as well as possible, the original mean, range, and standard deviation of the larger group in terms of their original n Ach scores. The stimuli, in final form, consisted of eight single type-written pages lettered A through H with one page for each subject and four short TAT stories per page. The order of the four stories corresponded to the order of presentation of the stimulus cards which was B, A, H, G in this case (see Appendix).

The Judges

The four judges, students at Michigan State University, represented, in a loosely defined manner, points on a continuum of increasing experience with projective materials. The most "naive" judge was an undergraduate Art-Social Science major; the next judge was a Psychology graduate student having no experience with projectives; the third, a Psychology graduate student with "some" experience; and the last, a fourth year Psychology graduate student with "much" experience. The first judge was female and the remaining three were males.

The Rating Instructions

Each judge was given an instruction packet (Appendix) which contained the following items: a description of how the TAT was administered and the stories obtained, procedural instructions for making the ratings, a list of rating criteria, three example and three practice protocol-sets, and practice rating sheets. The rating criteria consisted of a two-page

condensation of the n Ach scoring categories contained in the McClelland et al. (1958) scoring manual. All references to any kind of numerical scoring system as well as most of the elaborations and example materials were eliminated. In other words, the judge was not to learn the McClelland scoring system, but to use the defined categories (criteria) as a general base or guideline for determining his overall impression as to the relative magnitude of need achievement reflected in one protocol-set when compared with another.

The procedure was as follows. The judge was asked to read or reread each of the two protocol-sets for each comparison. For each subject pair, a labeled sheet of paper with a 20 cm. line had been provided. The first decision was to judge, according to the criteria, which member seemed to have the most need for achievement and label the paper accordingly. Letting the entire length of the line represent the total magnitude of that member's need for achievement, the second decision was to judge how much the other member seemed to have compared to the first and place a mark on the line to represent the relative magnitude (measured from the left side of the line). The judge was instructed to make his ratings as independent as possible by not referring back to previous ratings. For eight subjects, there were a total of 28 separate comparisons to be made after eliminating duplications such as A-B, B-A and self-comparisons such as A-A, B-B.

To help maximize the independence of comparisons, the order for presentation of the stimulus pairs was determined by a mathematical procedure developed by Ross (1934) for calculating "optimal" orders. "Optimal" means that (a) the order is "balanced" i.e. a given stimulus occurs an equal number of times as the first and second member of a pair (provided

that n is odd) and (b) pairs having the same member in common are maximally separated in their order of presentation. Slight adjustments, having little significance, must be made when the number of stimuli is even. (The order used is contained in the Appendix.)

For further study and clarification, example stories given by three subjects to the same stimulus card series had been selected from the 30 illustrative stories in McClelland et al. (1953, Appendix I). The protocols were intended to exemplify "very high" (n Ach = +21), "moderate" (n Ach = +10) and "very low" (n Ach = -2) need achievement stories respectively. Components of the stories relevant to the criteria had been underlined. In addition, three practice protocol-sets selected from the unused protocols in Klinger's achievement-aroused group were provided. The judges rated these stimuli on practice rating sheets.

After careful study of the instruction materials and completion of the practice ratings, the judges met with the experimenter to ask questions and discuss their practice ratings. At this time they were given the eight sets of stories and 28 labeled rating sheets arranged in the pre-determined order.

Derivation of the Scale Values

For reasons already given, the strategy of this study had dictated the desirability of employing a technique whereby scale values can be derived from unrepeated ratings of a single judge. Thus, the discussion to follow always refers to single ratio observations rather than to averaged observations (over N judges or N replications per judge). Computation of the scale values will be repeated for each judge.

For each comparison, the distance from the beginning of the line

to the mark is divided by the entire length of the 20 cm. line to form a ratio of the perceived magnitude of the smaller stimulus relative to the larger. Next, an n x n ratio matrix of all possible paired comparisons is formed. Half of the values (excluding the principal diagonal) consist of the directly observed ratios and are, by definition, less than one. The principal diagonal values are equal to one, and the remaining cell entries are obtained by taking the reciprocals of the observed ratios. Table 1 is a hypothetical matrix showing all possible comparisons of three stimuli A, B, and C. Each cell entry is the value obtained from dividing the column stimulus by the row stimulus.

Table 1. A Ratio Matrix of All Possible Paired-Comparisons of Three Stimuli

	A	B	C
A	$\frac{A}{A}$	$\frac{B}{A}$	$\frac{C}{A}$
B	$\frac{A}{B}$	$\frac{B}{B}$	$\frac{C}{B}$
C	$\frac{A}{C}$	$\frac{B}{C}$	$\frac{C}{C}$

Reverting, now, to a more general notation, consider Matrix R' (Table 2). Each R'_{jk} entry represents the observed column stimulus S'_k divided by the row stimulus S'_j. If S_k denotes the <u>true</u> scale value of stimulus k and R_{jk} the <u>true</u> ratio, then

$$(1) \qquad R'_{jk} = \frac{S'_k}{S'_j}$$

is an estimate of the ratio $R_{jk} = \frac{S_k}{S_j}$.

Also, we have

$$R'_{jk} = \frac{1}{R'_{kj}} \qquad (2)$$

that is, the entry for the kth column and jth row is the reciprocal of that in the jth column and the kth row, the subscripts being interchangeable since the column and row stimuli are the same.

Table 2. Matrix R' with Entries R'_{jk} Denoting the Ratio of the kth Column Stimulus to the jth Row Stimulus.

	1	2	. .	k	. .	n
1	R'_{11}	R'_{12}	. .	R'_{1k}	. .	R'_{1n}
2	R'_{21}	R'_{22}	. .	R'_{2k}	. .	R'_{2n}
.	.	.		.		.
j	R'_{j1}	R'_{j2}	. .	R'_{jk}	. .	R'_{jn}
.	.	.		.		.
n	R'_{n1}	R'_{n2}	. .	R'_{nk}	. .	R'_{nn}

If Comrey's analytical procedure were followed at this point with <u>these</u> data, we would begin by averaging R'_{jk} over j for each k to obtain an estimate of the rank order of the stimuli and then construct a new matrix with the rows and columns ordered with respect to increasing magnitude of the stimuli. He then makes use of the fact that, in addition to the direct estimate of the ratio S_k/S_j, n - 2 additional estimates of the same ratio can be obtained from the data matrix. For example, the ratio S_1/S_2 can,

in addition to the direct estimate, be estimated by

$$\frac{R'_{31}}{R'_{32}} = \frac{\frac{S'_1}{S'_3}}{\frac{S'_2}{S'_3}} = \frac{S'_1}{S'_2}$$

and similarly by R'_{41}/R'_{42}, R'_{51}/R'_{52}, etc. In this manner, Comrey finds n - 1 estimates of all adjacent pairs of stimuli: S_2/S_3, S_3/S_4 ... S_{n-1}/S_n, and computes an arithmetic average of these n - 1 estimates of each ratio ($\bar{R}_{jk} = S''_k/S''_j$). By arbitrarily assigning some scale value to the first stimulus (e.g. $S''_1 = 1$) the remaining scale values can be obtained directly. $\bar{R}_{12} = S''_2/S''_1$ so $\bar{R}_{12} = S''_2$. Having found S''_2, we have $S''_3 = S''_2\bar{R}_{23}$ and so on. Torgerson (1958, p. 108) argues that the use of the <u>arithmetic</u> mean in this case leads to inconsistencies because the mean of a set of ratios is not equivalent to the reciprocal of the harmonic mean of those ratios. Consider the hypothetical ratios 1, m, n, p. Then, in general

$$\frac{1+m+n+p}{4} \neq \frac{1}{\frac{\frac{1}{1}+\frac{1}{m}+\frac{1}{n}+\frac{1}{p}}{4}}$$

The problem resulting from this inequality becomes apparent when we consider that an average across rows (or columns) involves some entries that are reciprocals of the directly observed ratios. On the other hand, the <u>geometric</u> mean of a set of ratios, the nth root of the product of n ratios, <u>is</u> equal to the reciprocal of the geometric mean of the reciprocals. That is

$$(1 \cdot m \cdot n \cdot p)^{\frac{1}{4}} = \frac{1}{\left(\frac{1}{1} \cdot \frac{1}{m} \cdot \frac{1}{n} \cdot \frac{1}{p}\right)^{\frac{1}{4}}}$$

Thus, the geometric mean would seem to be a better average in this case.

If the ratios or root equations were converted to logarithms, we find that (a) geometric means are easier to compute since the log of the geometric mean of a set of ratios is simply the average of the logs of the ratios--that is:

$\log (1 \cdot m \cdot n \cdot p)^{\frac{1}{4}} = 1/4 \log (1 \cdot m \cdot n \cdot p) = 1/4(\log 1 + \log m + \log n + \log p)$,

(b) the use of logarithms converts the original multiplicative equations to additive ones and thus permits a solution of the scale values based on the principle of least-squares, (c) these additive properties will also permit an analysis of variance type development of a test for internal consistency.

With minor elaborations, the following derivation of the least-squares solution is parallel to that given by Torgerson (1958, pp. 109-111) for Comrey's data. That this is so derives from the fact that the solution is independent of both the method of obtaining the observed ratios and the number of observations per cell.

Converting the basic equation (1) to logs, we have the observed

$$\log R'_{jk} = \log S'_k - \log S'_j \tag{3}$$

as estimates of the true $\log R_{jk}$. Let us assume for the moment that from our matrix of observed ratios we have obtained the final <u>derived</u> estimates ($\log S''_k$) of the logs of the true scale values ($\log S_k$). Let $\log R''_{jk}$ denote the difference between any pair of derived estimates i.e. $\log R''_{jk} = \log S''_k - \log S''_j$. For each cell entry, $\log R'_{jk}$ we could then compute a <u>corresponding</u> $\log R''_{jk}$ from the derived logs of the scale values. If the data were errorless, $\log R''_{jk}$ would be exactly equal to $\log R'_{jk}$; but, since they are not, a difference or error ($\log R'_{jk} - \log R''_{jk}$) will exist. Our problem will be to find values of $\log S''_k$ from the observed

matrix of $\log R'_{jk}$ values for which the sum of squares of these differences is a minimum. That is, we want to minimize Q where

$$Q = \sum_{j}^{n}\sum_{k}^{n} (\log R'_{jk} - \log R''_{jk})^2 . \tag{4}$$

Note that since $\log R''_{jk} = \log S''_{k} - \log S''_{j}$, we have

$$Q = \sum_{j}^{n}\sum_{k}^{n} (\log R'_{jk} - \log S''_{jk} + \log S''_{j})^2 . \tag{5}$$

To simplify the derivation, let

$$x'_{jk} = \log R'_{jk}$$

$$y''_{k} = \log S''_{k}$$

$$y''_{j} = \log S''_{j}$$

Now,

$$Q = \sum_{j}^{n}\sum_{k}^{n} (x'_{jk} - y''_{k} + y''_{j})^2 . \tag{6}$$

We know from equation (2) that any entry in the original ratio matrix for the kth column and jth row has a corresponding reciprocal on the opposite side of the principal diagonal. In the log matrix, however, the corresponding entry in the jth column and kth row is simply the log of the original ratio with a reversal in sign i.e. $\log 1/R_{jk} = -\log R_{jk}$. From this principle we can now observe that the sum of entries in the kth column is equal to the sum of entries in the corresponding kth row except for a reversal in sign. The same obviously holds for a matrix of all possible $(x'_{jk} - y''_{k} + y''_{j})^2$ values. Hence, we may be concerned only with the columns i.e. the summation across rows. Equation (6) becomes

$$Q = \sum_{j}^{n} (x'_{jk} - y''_{k} + y''_{j})^2 . \tag{7}$$

To find the y''_{k} values which minimize Q, we now take the partial derivative of Q with respect to y''_{k}, set it equal to zero, and solve for y''_{k}.

$$\frac{\partial Q}{\partial y_k''} = 2 \sum_j^n (x_{jk}' - y_k'' + y_j'')(-1) = 0$$

Dividing by 2 and rearranging, we get

$$\sum_j^n y_k'' = \sum_j^n x_{jk}' + \sum_j^n y_j''$$

which gives $$ny_k'' = \sum_j^n x_{jk}' + \sum_j^n y_j''$$

Solving for y_k'' we obtain

$$y_k'' = \frac{1}{n} \sum_j^n x_{jk}' + \frac{1}{n} \sum_j^n y_j' \tag{8}$$

It is possible to express y_k'' solely as a function of x_{jk}' by specifying a unit of measurement such that the geometric mean of the derived scale values is equal to one. It follows that the average of the logs of the derived scale values is equal to zero, i.e. $\frac{1}{n} \sum_j^n y_j'' = 0$. Equation (8) then reduces to

$$y_k'' = \frac{1}{n} \sum_j^n x_{jk}' \tag{9}$$

Converting back to the original notation, we have

$$\log S_k'' = \frac{1}{n} \sum_j^n \log R_{jk}' \qquad (k = 1, 2, 3, \ldots, n) \tag{10}$$

Hence, for any value of k, the logs of the derived scale values are found by averaging the logs of the observed ratios in each column. The final scale values are found by taking the antilogs of log S_k''.

A Test for Internal Consistency

In the preceding section we found that for each log R_{jk}' in the log matrix of observed ratios, we could compute a corresponding log R_{jk}'' from

the derived scale values. Implicit in the averaging procedure for obtaining the derived scale values was a basic assumption of the model i.e. that observed ratios between stimulus pairs are linearly consistent with other ratios in the same matrix. Furthermore, it is assumed that the stimuli are assigned final scale values which represent points on a linear or unidimensional conttinuum of the attribute in question. One measure of the validity of these assumptions is the degree to which ratios calculated from the n derived scale values agree with or predict the directly observed ratios. The present task will be to develop an overall measure of the "goodness of fit" between the observed ratios ($\log R'_{jk}$) and the derived or predicted $\log R''_{jk}$ values in terms of the proportion of total variance in the observed data that can be accounted for by the theory. This measure will <u>also</u> be interpreted as an assessment of intra-judge reliability or the degree to which a judge has been internally consistent with his own ratings.

As the original data matrix has been converted to logs, the cell entries can now be expressed as differences between stimuli instead of ratios. This feature now permits an analysis of components of variance approach[1] similar to the development by Gulliksen and Tukey (1958) for Thurstone's Case V model. The developments differ, however, in that Thurstone's model (a) employs N ordinal-type observations per cell and requires an arc sine transform of the observed proportions due to binomial sampling variance, and (b) has a "known" binomial error component which makes possible a separate estimate of deviation from linearity.

---

[1]The standard linear model from classical measurement theory is $X = t + e$ where X is an observed value, t the hypothetically <u>true</u> value and e the error component. As the original data here was in the form of ratios, the above equation would have been written $X = t \cdot e$ where error is correlated with the magnitude of the ratio judgment. However, in log form, $\log X = \log t + \log e$, the data may be treated in the customary manner.

Expanding and summarizing our notation of the preceding section, we have the following definitions in log form:

$y_k = \log S_k$ (true scale value)

$y_j = \log S_j$ (true scale value)

$x'_{jk} = \log R'_{jk} = \log S'_k - \log S'_j$ (observed ratio)

$x''_{jk} = \log R''_{jk} = \log S''_k - \log S''_j$ (derived ratio)

Each observed cell entry x'_{jk} will be considered to consist of three components or contributions. The first is the difference between the true scale values, $y_k - y_j$, with each y drawn from a population with variance σ_y^2. Thus, the variance of the differences is considered to be drawn from a population with variance $2\sigma_y^2$. The second is a deviations component (d') due to deviations from the linear assumptions of the model with variance $\sigma_{d'}^2$, and finally, an error component (e) drawn from a population with variance σ_e^2. For our model, which employs only one ratio observation per cell, an independent estimate of the error variance is not available. Hence, for the remainder of the discussion, there will be little point to referring to the last two components as separate entities since it is not possible to discriminate between them computationally. We shall define, then, a residual component (d) consisting of the deviation and error components combined, and drawn from a population with variance σ_d^2 and an assumed $E(d_{jk}) = 0$. Our basic linear equation may then be written as

$$(1) \qquad x'_{jk} - \bar{x}' = (y_k - y_j) + d_{jk}$$

with population variances $2\sigma_y^2$ and σ_d^2 respectively.

To obtain our mean square estimates of these variance components, we shall begin by partitioning the observations into their component sums

of squares. We have

$$(2) \qquad x'_{jk} - \bar{x}' = (x''_{jk} - \bar{x}') + (x'_{jk} - x''_{jk})$$

where the deviation of an observed cell entry from the grand mean of the matrix is expressed as the sum of a linear component $(x''_{jk} - \bar{x}')$ and a residual component $(x'_{jk} - x''_{jk})$. Because of the symmetry of the matrix of x'_{jk} values above and below the principal diagonal, $\sum_j^n \sum_k^n x'_{jk} = 0$ and hence $\bar{x}'_{jk} = 0$. Equation (2) may now be written as

$$(3) \qquad x'_{jk} = x''_{jk} + (x'_{jk} - x''_{jk}).$$

Summing and squaring,

$$\sum_j^n \sum_k^n x'^2_{jk} = \sum_j^n \sum_k^n x''^2_{jk} + 2 \sum_j^n \sum_k^n x''_{jk}(x'_{jk} - x''_{jk}) + \sum_j^n \sum_k^n (x'_{jk} - x''_{jk})^2$$

We may now define the three sums of squares.

Total SS $\qquad S_T = \tfrac{1}{2} \sum_j^n \sum_k^n x'^2_{jk}$

Linear SS $\qquad S_L = \tfrac{1}{2} \sum_j^n \sum_k^n x''^2_{jk}$

Residual SS $\qquad S_D = \tfrac{1}{2} \sum_j^n \sum_k^n (x'_{jk} - x''_{jk})^2$

The term $(\tfrac{1}{2})$ is used because the data matrix is based on only $n(n-1)/2$ independent observations. (The diagonal values are, by definition, equal to zero.) The cross-product term, $2 \sum_j^n \sum_k^n x''_{jk}(x'_{jk} - x''_{jk})$ does not vanish, nor is this a necessary condition. If S_D computed directly by $\tfrac{1}{2} \sum_j^n \sum_k^n (x'_{jk} - x''_{jk})$ turns out to be substantially the same as $S_T - S_L$ computed by $\tfrac{1}{2} \sum_j^n \sum_k^n x'^2_{jk} - \tfrac{1}{2} \sum_j^n \sum_k^n x''^2_{jk}$, then the contribution of the cross-product term may be disregarded. The justification of this deletion was verified by the empirical data from the present study[1].

---

[1]That is, using the alternate computational procedures described, the difference between the final variance ratios were, on the average, 0.002. Hence, for all practical purposes we may consider the linear and residual components to be essentially orthogonal. This assumption may possibly be invalid for other methods of deriving the scale values.

The degree of freedom for Total, Linear, and Residual sums of squares are $n(n-1)/2$, $(n-1)$, and $(n-1)(n-2)/2$ respectively. We may now compute our mean square estimates of the population variances.

Total mean square $\quad T = \dfrac{S_T}{\frac{n(n-1)}{2}} \quad$ where $E(T) = 2\sigma_y^2 + \sigma_d^2$

Linear mean square $\quad L = \dfrac{S_L}{n-1} \quad$ where $E(L) = n\sigma_y^2 + \sigma_d^2$

Residual mean square $\quad D = \dfrac{S_D}{\frac{(n-1)(n-2)}{2}} \quad$ where $E(D) = \sigma_d^2$

In accordance with the standard definitions, the coefficient of linear consistency or variance ratio shall be defined as the ratio of the true or linear variance (the variance accounted for by the derived scale values) to the total variance of the direct observations, or

$$(4) \qquad \frac{2\sigma_y^2}{2\sigma_y^2 + \sigma_d^2} = \rho_{yy}$$

Our estimate of this ratio may be found by

$$(5) \qquad r_{yy} = \frac{T - D}{T} \doteq \frac{2\sigma_y^2}{2\sigma_y^2 + \sigma_d^2} = \rho_{yy}$$

This coefficient, r_{yy}, then, is the overall measure of "goodness of fit" between the logs of the observed ratios and the logs of the ratios computed from the derived scale values, and is expressed in terms of the proportion of total variance that can be accounted for by the linearity assumptions of the model.

To briefly summarize the computational procedures, we first find the sum of squared log R'_{jk} from the log matrix of observed ratios using only the values above, or equivalently, below the principal diagonal. Dividing by

the number of independent observations, n(n-1)/2, we obtain the Total mean square estimate T. Substracting the half-matrix of derived log R''_{jk} values from the corresponding half-matrix of observed log R'_{jk} values, a half matrix of residual values is obtained. The sum of squared residuals divided by (n-1)(n-2)/2 gives the residual mean square estimate D. r_{yy} is then obtained from equation (5).

For those who would be interested in a more familiar computational procedure and would settle for a description of the data from a <u>particular</u> sample, it should be noted that the square of the Pearson r between the observed x'_{jk} and the derived x''_{jk} values is a very close approximation to the above measure provided that the linearity assumption is tenable. From the four data matrices of the present study, r_{yy} was found to be only a slightly more conservative estimate than the Pearson r squared, an average difference of about 0.03.

RESULTS

When logs of the scale values, as opposed to their anti-logs, were used for the data analyses, it was found that (a) greater homogeneity of variance was maintained between the two treatment groups, and (b) inter-correlations between raters, and between raters and the McClelland scores, were higher. This latter observation agrees with the findings of O'Grady and Karon (1968). These authors noted, from inspections of scatter plots, that their raters' log scale values appeared to show more linearity in the relationship between raters and between raters and an external criterion than the anti-logs. Both (a) and (b) above may be explained in part by the fact that logarithms tend to reduce or "pull-in" the relatively extreme values resulting from judgments of very high similarity or dissimilarity between stimuli. Considering the above, the following data anlyses were carried out with the logs of the derived scale values.

Table 3. Transformed Log Scale Values and McClelland Scores for the Eight Sets of TAT Stimuli

		Judges				McClelland scores
	Stimulus	I	II	III	IV	
n Ach Group	D	8.73	16.20	13.93	13.18	14.00
	A	12.84	8.61	9.08	11.22	9.00
	F	4.44	5.85	7.14	5.74	10.00
	B	1.06	2.17	3.21	3.14	3.00
Neu Group	H	7.57	5.31	6.62	8.27	8.00
	C	11.75	5.76	7.04	7.50	5.00
	G	5.29	7.09	6.62	3.66	3.00
	E	-1.67	-0.98	-3.63	-2.73	-2.00

The scale values of the eight stimuli for each of the four judges are reported in Table 3 along with the original McClelland scores. For visual inspection purposes only, the original values were converted to z scores to obtain equal units of measurement between judges, and then again converted by the linear transform, $X' = 4.6837\ (z_x) + 6.25$[1], so that their magnitude and separation could be compared directly with the McClelland scores.

Table 4. The Internal Consistency (Intra-Judge Reliability) Coefficients (r_{yy}) for Four Judges.

	Judges			
	I	II	III	IV
r_{yy}	0.91	0.93	0.97	0.93

Table 4 shows the internal consistency (intra-judge reliability) measures obtained for the four judges. Between 91 and 97 percent of the total variance in ratio observations could be accounted for by the derived scale values, which in turn are based on the linearity assumptions of the model. The magnitude of these figures indicates that the data did, in fact, fit the model, and that the judges were able to make consistent ratings with the materials provided.

The inter-judge reliabilities and concurrent validity coefficients are found in Tables 5 and 6 respectively. With the exception of Judge I (the least experienced), the judges agreed quite well with each other and with the expert McClelland scorer. When queried after the ratings were completed, Judge I gave some indications that she did not attempt to

[1]The 4.6837 and 6.25 are the standard deviation and mean respectively of the eight McClelland scores.

Table 5. Inter-Judge Reliabilities for the Four Judges (I - IV)

	I	II	III
II	0.65		
III	0.77*	0.95*	
IV	0.85*	0.88*	0.94*

* $p < 0.05$, 2-tailed

identify specific achievement goals within sets of stories. In addition, she communicated to the experimenter some inclinations to either "like" or "dislike" the subjects on the basis of their written protocols. From an examination of Table 3 (and also Table 7) it seems apparent that she used somewhat different criteria than the other raters even though there appears to be some general overall agreement. It is interesting to note that, using somewhat modified criteria with the same set of instructions and materials, she was nevertheless consistent with her own ratings.

Table 6. Correlations between Log Scale Values of the Four Judges with the McClelland Scores.

Judges			
I	II	III	IV
0.63	0.86*	0.91*	0.91*

* $p < 0.05$, 2-tailed

Table 7 presents a summary analysis of the measurement precision of the scale values. Again, numbers are in converted form for ease of

Table 7. Summary Analysis of the Log Scale Values versus the McClelland Scores Based on Comparisons between the Treatment Groups.

	Judges				McClelland
	I	II	III	IV	scores
n Ach (n=4)					
M	6.77	8.21	8.34	8.32	9.00
SD	5.12	5.94	4.46	4.67	4.55
Neu (n=4)					
M	5.74	4.30	4.16	4.18	3.50
SD	5.61	3.60	5.20	5.03	4.20
t*	0.27	1.13	1.22	1.21	1.77
est. ω^2	0.00	0.03	0.06	0.06	0.21

* $t_{.05} = 1.94$, 1-tailed, 6 df

comparison with Klinger's data. Differences between means were in the expected direction for all raters, but in no case was the separation as great as that afforded by the original McClelland scores. Dispersion of scale values within the groups was essentially equivalent. No t ratios were significant, although the non-significant t for Klinger's data was due to the loss in degrees of freedom by sub-sampling from the larger (n=10) n Ach group. However, the statistic of greatest interest here is the est. ω^2 (Hays, 1963, p. 327) which is an estimate of the proportion of variance in the scale values attributable to the difference between the groups. As such, it can be given an interpretation similar to the more familiar correlation ratio. We see that the est. ω^2 values based on the new scales ranged from 0.00 to 0.06, while that for the McClelland scores was 0.21. The logic of the experimental design in conjunction with the substantially lower est. ω^2 for the new scale values indicates that improvement in measurement precision, as here defined, has not been demonstrated.

DISCUSSION

The study, as well as the two by O'Grady and Karon (1968) and Carter (1966), offers a rather substantial demonstration of the scalability of projective responses which have been presently defined as stimuli. It establishes the tenability, at least, of the following inferences: the relatively simple personality variable, need Achievement, can be treated as a unidimensional construct; sophisticated judges using a TAT measure of that construct are able to make the most difficult of comparisons between stimulus pairs i.e. direct ratio estimates; and finally, these ratio estimates can be manipulated in such a way as to assign numbers to stimuli which represent points on a ratio scale of measurement as operationally defined by the model. The reader who feels that the last inference is "stretched" a bit from the single measure, r_{yy}, does so with some justification. The internal consistency coefficient, which assesses the degree of independence between the final scale values and any particular ratio from which they were derived, meets but one of three criteria for demonstrating the ratio properties of a scale.

A second criterion is that the scale value of any stimulus can be determined to within a linear transformation of the form $y=ax$ (the absolute zero requirement). An appropriate test of this property would be a demonstration of the independence of the scale of the particular set of stimuli used in its construction. To do this in the classical manner requires the presence of an external referrent against which a linear plot

could be made with points on the scale. As such a referrent has not been defined for the attribute of present concern, an experimental verification of the second criterion seems to depend on the possibility of finding a substitute procedure for demonstration of the independence property. One plausible substitute would be to show the independence of the ratio of any two stimuli of the stimulus context from which it was originally derived. For example, one could rescale stimuli A and B with six new stimuli and then determine whether the new ratio agrees with the previous one. Considering that it is really the protocols that are the stimuli, the absolute zero referrent could be defined as any set of TAT stories with "totally unrelated achievement imagery" (as defined by McClelland).

The third criterion for establishing ratio properties of the scale is, of course, the requirement that equal differences in scale values correspond to subjectively equal intervals. With a larger number of stimuli (than used for this study) some methods of "equal appearing intervals" are quite suitable for demonstrating this property and are easily adaptable to the characteristics of the stimuli presently used. In summary, the point to be made here is that the r_{yy} measure satisfies but one of three conditions required for final verification of the ratio properties attributed to the scale, and that additional experimentation is required for the remaining two.

Having discussed at some length the supposed mathematical properties of the scale values, what now can be said of their validity, i.e. the extent to which they afford prediction of the external criterion, treatment conditions. We have seen that, with one exception, the judges

agreed quite well with each other and with the McClelland score values, and yet, the separation between treatment group means afforded by the scale values was less than for McClelland's scores. The extremely small sample may, of course, be considered a negative factor in making such a comparison. Unfortunately, this was necessitated by the number of comparisons required to fill a complete data matrix. For future work, an alternate procedure is available for finding "iterative" solutions of the scale values from a partial data matrix of only 2(n-1) comparisons. This would permit an increase to at least 20 or 30 stimuli before the scaling task would start to become unreasonable.

A second difficulty in obtaining better validity coefficients may have arisen from the fact that the judges were required to cumulate their impressions over four stories for each subject. It would seem that this procedure may contain a greater error liability than would have been the case if the judge were asked to scale the eight stimuli for each story separately. In the latter case, it would be feasible to mathematically combine the scale values for each story, and this would obviate the necessity for the rater to do it in his head by some unknown process. This is after all closer in concept to the way a single McClelland score is obtained for n Ach. A further study is planned to investigate whether this procedure would result in an increase in predictive validity of the scale values.

Finally, the original experimental sample was small enough in size as to make it impossible to eliminate protocol sets that were "very high" or "very low" in need achievement imagery. When two or three stimuli of moderate magnitude are each compared with a very high or low magnitude

stimulus, those comparisons act as relatively poor discriminators between the stimuli of moderate magnitude. A larger sample would permit a better selection of stimuli.

In assessing the overall usefulness of the technique of personality measurement investigated by this study, we must recognize the nature of its specific limitations as well as its specific advantages. We observe, for example, that a group of subjects is required for assignment of a magnitude estimate to a particular individual. Hence, the technique must be considered totally inappropriate for many types of clinical assessment settings. Furthermore, until the absolute zero requirement has been satisfied by experimentally demonstrating independence between scale values and the particular stimuli used in their construction, it would not be possible to compare scale values across samples. That is, the scale value assigned to a particular individual depends on his standing relative to other individuals in the sample scaled. Correspondingly, the measure assigned to John Doe at Michigan State University could not be compared with the measure assigned to Pete Smith at the University of Southern California. However, the potential usefulness of this type of measure in a research setting seems more promising. Pending that further experimental improvements of the kind suggested above should prove fruitful, a scale of the type developed here could be used as a powerful criterion measure for the development of simpler but accurate assessment devices. If measurement theory does, in fact, have some connection with reality, then we ought to predict that our method of assigning numbers to the magnitude of personality attributes will open the door to higher levels of measurement precision and ultimately the mathematical

expression of functional relationships between personality variables. It goes without saying that such glorified notions are yet to be demonstrated.

BIBLIOGRAPHY

BIBLIOGRAPHY

Carter, R. E. The measurement of emotional health through the use of Estavan's modified paired comparison technique. Unpublished master's thesis, Michigan State University, 1966.

Comrey, A. L. A proposed method for absolute ratio scaling. Psychometrika, 1950, 15, 317-325.

Gulliksen, H., & Tukey, J. W. Reliability for the law of comparative judgment. Psychometrika, 1958, 23, 95-110.

Haber, R. N. The prediction of achievement behavior by an interaction of achievement motivation and achievement stress. Unpublished doctoral dissertation, Stanford University, 1957.

Hays, W. L. Statistics for psychologists. Holt, Rinehart, & Winston, 1963.

Klinger, E. Modeling effects on achievement imagery. Journal of Personality and Social Psychology, 1967, 7, 49-62.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. The Achievement Motive. New York: Appleton-Century-Crofts, 1953.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. A scoring manual for the achievement motive. In J. W. Atkinson (Ed.) Motives in fantasy, action, and society. Princeton, N. J.: Van Nostrand, 1958, 179-204.

McClelland, D. C., Clark, R. A., Roby, T., & Atkinson, J. W. The projective expression of needs: IV. The effect of the need for achievement on thematic apperception. Journal of Experimental Psychology, 1949, 39, 242-255.

O'Grady, P. & Karon, B. P. Quantified judgments of mental health from the Rorshach, TAT, and clinical status interview by means of a scaling technique. Unpublished paper, Michigan State University, 1968.

Ross, R. T. Optimum orders for the presentation of pairs in the method of paired comparisons. Journal of Educational Psychology, 25, 375-382.

Torgerson, W.S. Theory and methods of scaling. New York: Wiley, 1958.

APPENDIX

APPENDIX

SAMPLE INSTRUCTION PACKET USED BY THE JUDGES

INTRODUCTION

This is an experiment in ratio scaling, the purpose of which is to determine how accurately and reliably projective materials like the TAT can be scaled with the model to be presently used. The materials to be used are eight sets of short TAT stories with four stories per set. The eight sets of stories have been written by eight subjects who presumably differ from each other along a psychological continuum defined as the "need for achievement." The fantasy content of these stories is presumed to reflect the amount of the need for achievement possessed by the storyteller. Your specific task will be to rate these subjects in pairs by estimating how much need achievement one person seems to have when compared with another. From your ratings, a final scale value will be derived for each of the eight subjects which should reflect the person's relative position on the need achievement continuum.

The TAT Stories

The stories you are to rate were obtained in the following manner. Four 8½ x 14" sheets of paper were given to each subject. Four sets of printed questions were equally spaced on each sheet. The questions were intended to elicit complete coverage of a plot. They were:

(1) What is happening? Who are the persons?

(2) What has led up to the situation? That is, what has happened in the past?

(3) What is being thought? What is wanted? By whom?

(4) What will happen? What will be done?

Each of four pictures was presented to the subjects for 20 seconds. Immediately after viewing each picture, the subject was given four minutes to write his answers to the above questions, one minute for each question. The four stories written by each subject have been reproduced as written (no attempt being made to alter grammar, spelling, or punctuation) on a single sheet of paper, one sheet for each subject. Most of the stories contain four short paragraphs, each corresponding to one of the above four questions. As the four questions have not been reproduced with the stories, you may refer back to them any time you wish.

The pictures corresponding to each of the four stories are as follows:

Story 1 -- Picture B -- Two men ("inventors") in a shop working at a machine.

Story 2 -- Picture A -- "Father-son." Card 7 BM from the Murray TAT.

Story 3 -- Picture H -- Boy in checked shirt at a desk, an open book in front.

Story 4 -- Picture G -- Boy with vague operation scene in background. Card 8 BM.

A copy of each picture is included in this instruction packet.[1]

---

[1]These pictures are not included in this Appendix. Pictures A and G are from the standard TAT series. Pictures B and H are in McClelland et al. (1953, pp. 100-101).

METHOD

As mentioned above, you will be rating the eight story sets in pairs. The story sets for the eight subjects have been lettered A through H. This scaling model rquires that each subject be compared with every other subject one time. Hence, when we eliminate duplications such as AB, BA and self comparisons such as AA and BB, we are left with a total of 28 pairs to be rated.

Take one pair at a time <u>according to the order in which they are listed on the last page of these instructions</u>. Read each set of stories for the pair to be rated. The first set to be read is the first one listed in the pair. Example: for pair A-D, read set A first, then D.

<u>Your first decision</u> will be to judge, according to the criteria below, which story set seems to represent the person who has the most need for achievement of the pair. For each pair, a sheet of paper with a 20 cm. line has been provided. Label this paper in the space provided at the upper right with the letter of the subject that you have judged to have the most need for achievement. You now let the entire length of the line represent the total magnitude or amount of need for achievement that member has.

<u>Your second decision</u> will be to judge how much need achievement the other member of the pair seems to have compared with the first and place a mark on the line which represents this relative amount. For example: you are considering the hypothetical pair M-T and you decide that T has more need for achievement than M. You label the sheet T and let the line equal the total amount of need for achievement that T seems to have. In comparison to T, you feel that M has about two-thirds as much achievement motive. You then place a mark approximately two-thirds of the distance from the left end of the line. If you feel that M has about half as much, then place the mark in the middle of the line, and so forth.

Follow the exact same procedure for rating each of the other pairs making sure that you read both sets of stories for each comparison. This process will, of course, require rereading some of the same stories over again for the different comparisons.

As you realize by now, your painstaking care in making each decision is most critical to the validity of this procedure. You may, at the time you are judging, wish to reread parts of one or both sets of stories, part of the instructions, or parts of the criteria listed below. Feel free to do so. Try not to work when you are tired, and divide the task into separate time intervals. In other words, work at your leisure.

CRITERIA

General Definition of Need Achievement -- the desire to attain success in competition with some standard of excellence.

The follwing categories are to serve as a general guide for determining the amount of achievement motive (need, drive) reflected in a given set of stories. On the first reading you may find that you disagree with some of the categories or find that others seem to be irrelevant or unimportant. However, it is important to realize at the outset that these criteria were derived from a particular definition of the achievement motive, a particular theory of motivation, and certain empirical studies that have demonstrated their capability to discriminate between various subjects. It is these criteria, then, which define the variable we are interested in scaling, and hence, they should carry as much weight as possible when you make your final ratings.

The criteria are listed here in the order in which they will most likely appear in the imaginative stories.

General Achievement Imagery

Some reference in the story to an achievement goal is the first criterion to be taken as evidence that the story contains need for achievement. More specifically, the goal of some individual in the story is to be successful in terms of competition with some standard of excellence.

This goal may be stated or implied in any of the following ways: (a) One of the characters in the story may be engaged in some competitive activity where winning or doing as well as or better than others is actually stated as a primary concern (e.g. "The boy wants to win an essay contest."); (b) The desire to win may not be explicitly stated, but (1) affective concern over goal attainment is expressed ("The boy wins the essay contest and feels proud."), (2) certain types of Instrumental Activity are indicated ("The boy is working very carefully on his essay."); (c) The standard of excellence may involve no competition with others but meeting self-imposed requirements of good performance (e.g. "The boy is studiously and carefully preparing his homework."); (d) One of the characters may be involved in some unique accomplishment which will mark him as a personal success (e.g. inventions, artistic creations, etc.); (e) One of the characters may be involved in attainment of a long-term achievement goal i.e. being a success in life, becoming a machinist, doctor, lawyer, etc.

It should be emphasized here that the outcome of a story in terms of final success or failure is completely irrelevant for our purposes. However, emotional concern, either positive or negative, in connection with the outcome is considered evidence of a need or striving to achieve.

Stated Need for Achievement

Someone in the story may state the desire to reach an achievement goal. Examples are: "He wants to finish the painting.", "He is determined to get a good mark."

Instrumental Activity

Overt or mental activity by one or more characters in the story indicating that something is being done about attaining an achievement goal is considered additional evidence of achievement motivation. The outcome of the instrumental activity may be successful, unsuccessful or unknown (e.g. "... the young man is working hard on the piano, studying long hours ... He may not do very well.")

Anticipatory Goal States (positive or negative)

Someone in the story may anticipate either goal attainment or frustration and failure or both. A positive Anticipatory Goal State is exemplified when someone is thinking about the success he will achieve, expects that the invention will work, dreams of himself as a great surgeon, etc. The Anticipatory Goal State is negative when someone is worried about failure, is concerned over the possibility that the invention will not work, expects the worst, or is wondering whether or not he will succeed. The occurrence of either category, positive or negative, is considered additional evidence of the need for achievement.

Obstacles or Blocks

The presence of obstacles to the progress of goal-directed activity is another element that may occur as a contributing part of an achievement-related story. The obstacle may be located within the individual (lack of confidence, inability to make decisions, or some past failure) or outside as part of the environment (lack of finances, too much competition, etc.).

Nurturant Press

Forces in the story, personal in source, which aid the character in the story who is engaged in an on-going achievement-related activity are called Nurturant Press. Someone aids, sympathizes with, or encourages the person in his striving for achievement. For example: "The experienced machinist is trying to straighten things out for the apprentice and is encouraging him."

Affective States (positive or negative)

Affective (emotional) states associated with goal attainment, active mastery or frustration of the achievement-directed activity are additional elements contributing to the strength of achievement motivation reflected in a particular story. For example, someone in the story may experience: (1) a positive affective state associated with active mastery or definite accomplishment ("He enjoys painting", "He is proud of his accomplishment",

"They are very satisfied with their invention"), or (2) a negative affective state associated with failure to attain an achievement goal ("He is disturbed over his inability", "He is despondent, mad, and sorry.") In either case (positive or negative), affect in connection with evaluation of performance is considered an additional contribution to the assessment.

Achievement Thema

Sometimes the Achievement Imagery can be elaborated in such a manner that it becomes the central plot or "thema" of the entire story. That is, striving for an achievement goal and eventual attainment of the goal may be the central plot. In this case, additional weighting is given to the overall story.

You may have noticed that the criteria in the above list are not completely independent of each other. An example of this is that all the criteria depend pretty much on the appearance of the first one, "Achievement Imagery," for any given story. Nevertheless, the appearance of each separate category is considered to make an independent contribution to the overall strength of the need for achievement reflected in a set of stories for a given person. In other words, the total number of the above criteria which appear in an entire set of stories contributes to the magnitude of the projected achievement motive.

When you make your comparison between one subject and another on the basis of the amount of need for achievement that seems to be reflected in their sets of stories, no attempt should be made to enumerate the above categories to obtain any kind of "score." Rather, the categories are intended to serve as a general guide to help you to identify the elements in a story that we are considering to be important. In other words, they should serve as a base for determining your overall impression as to the relative amount of achievement motive that one subject has compared with another.

Following are sets of stories given by three different subjects to the four picture cards used for this study. They will exemplify stories which reflect very high, moderate, and very low need for achievement respectively. Words or phrases considered relevant to the need for achievement have been underlined. The stories are recorded as actually written by three college students from another university.

Subject A Example

Story 1

"Two draftsmen are fashioning the vital part of a new invention. They have all the necessary tools and parts for it.

"These men are anxious to revolutionize the tool and die industry. They have spent months, even years working out this new tool -- first on paper, later practically.

"Each waits with bated breath for the final part to be finished so they can put together the new tool -- they want satisfaction, honor and financial reimbursement.

"The invention will be a failure. Discouraged, and financially bankrupt, one will commit suicide, the other will give up and go to Alaska to try his luck at farming."

Story 2

"The elderly gent has a scheme to pull a fast one on the market and he is putting his stooge wise to it. The old gent is a sharp Wall Street operator, and the young fellow is his flunky.

"Thru the years, by his suave manner and personality the elderly gent has built up a lot of confidence in brokers, bankers, etc., in the hope of someday fleecing the lot of them.

"The young chap is debating the possibilities of the scheme succeeding while the older gent wonders just how well the young fellow will perform. Both want money and will go the limit to get it.

"The scheme will work. The elderly gent will take the lion's share and hit Paris, Monte Carlo, St. Moritz, etc. The young chap will go on into another scheme and get tripped up, landing in jail for embezzlement and ... ."

Story 3

"This chap is doing some heavy meditating. He is a sophomore and has reached an intellectual crisis. He cannot make up his mind. He is troubled, worried.

"He is trying to reconcile the philosophies of Descartes and Thomas Aquinas -- and at his tender age of 18. He has read several books on philosophy and feels the weight of the world on his shoulders.

"He wants to present a clear-cut synthesis of these two conflicting philosophies, to satisfy his own ego and to gain academic recognition from his professor.

"He will screw himself up royally. Too inexperienced and uninformed, he has tackled too great a problem. He will give up in despair, go down to the Goodyear and drown his sorrows in a bucket of Piel's."

Story 4

"The boy is reflecting on a scene which he has read about. The doctors are barred from the profession and they take only criminals, etc., as patients. The patient has been badly wounded in a gun fight and is being operated on.

"The boy has read a lot about doctors. He visualizes them as good and bad. By some quirk he would like to become a doctor of this sort depicted, so that he could experiment on people.

"The boy wants to be a surgeon so he can cut people apart and see what makes them tick. He is a sadist at heart even so young. He especially wants to cut up beautiful girls after raping them.

"He will go out at the age of eighteen or so, waylay some 12 year old girl, rape her and hack her to pieces with a butcher knife. After that he will attack an old lady, but the cops will arrive just in time and in the pursuit

Subject B Example

Story 1

"Two men about to break the machine. They are 2 of 3 partners in a little machine shop. They were endeavoring to build up a bigger inventing company.

"Their third partner was killed by a bolt that flew from the machine. This partner was loved dearly by them and after countless endeavors to ... .

"They are thinking of their departed partner. All that is wanted is revenge at the machine that was invented. Their dream invention has turned out to be a source of great misery for them.

"They will break the machine, go on working in a small business and never again attempt to invent machines or expand their business."

Story 2

"Father and son. The son has just finished confessing his guilt in leading some girl astray. The father is a distinguished person and the son had been thought of as one who would not stoop to such a low act of seduction.

"The boy due to his father's social importance, had to be careful of his action in society. Finally he could keep up the pretense no longer and gave vent to his natural passion.

"The father is thinking of disgrace and finds disappointment in his son. The son feels lost -- almost in a daze due to his recent confession which he thought he'd never tell.

"A psychiatrist will be called in to help the fellow find himself and the father will get a polite 'bawling out' that it was mainly through his fault that the incident happened."

Story 3

"He is a student, very conscientious forced to go to school by his parents with great ambitions of becoming a lawyer. He is a 'grind' due to the fact that outside pressure is put on him. He has no girl friend.

"His fraternity brothers are having a house party with plenty of girls and fun. He has little money to ask a girl down, decides to study during the party.

"A girl or companion is wanted by this freshman. Life seems bitter and he wishes he could be like the other fellows having a good time.

"He will get very frustrated, decide that he has just got to get a girl sooner or later if his parents expect him to do a good scholastic job."

Story 4

"The boy has just seen a movie involving a medical drama. He is now thinking of the incident and has been inspired to become a great surgeon.

"Aside from the movie, he also has been under a serious operation which saved his life and is thankful to the doctors. He hopes he can do the same some day.

"How wonderful it will be when I'm a doctor. All that is wanted is the fulfillment of becoming an M.D.

"He will be in medical school but the strain is too much for him and he will attempt suicide but will be stopped and gradually adjust himself to a happy life."

Subject C Example

Story 1

"Two men are working at repairing a machine. One is doing the hammering while the other looks on, probably criticizing. Both are machinists.

"The machine has probably gone on the bum, and now needs to be fixed. The one hammering has probably been called to the task by the other.

"The man on the right wants the other man to fix his machine because something has gone wrong.

"The machine will be fixed so that the man on the right can go back to work."

Story 2

"Here we have father and son talking over some serious problem (probably married life).

"The son has probably been arguing with his wife, who suddenly decided to leave him, and now he resorts to his father for advice.

"The father is telling the son that he shouldn't be hasty about such things and to return to his wife and mend their argument.

"The son will probably go back to his wife and tell her that he is sorry for the way in which he had acted."

Story 3

"This poor boy is now worrying about something he has done recently. He is probably wondering why he did it.

"This boy has probably had his first sexual intercourse with a girl after which he was given a lecture on V.D.

"He is probably thinking about why he has done this 'awful deed' (so they say).

"He will probably say to himself that it was wrong to do what he did and that he won't do it any more."

Story 4

"The boy is thinking of some incident that has happened in a book which he just read. The persons other than the boy are characters of the book he has read.

"The boy has read a book in which the incidents are constantly playing on his mind.

"The boy is thinking of how these two men have just murdered another man for talking too much.

"The two men will continue in their scheme and bury their dead man, while the boy will probably have nightmares."

If you were actually comparing these subjects, you might have proceeded as follows: for the pair A-B, you first decided that A had more need for achievement than B. You labeled the paper with an A and your mark on the line indicated that B had say somewhere between one-third and two-thirds as much need achievement as A. For pair B-C, you decided that B was greater than C and that C had about one-forth to one-third as much as B. Finally, your comparison between A and C showed that A had more than C and that C had say one-tenth as much as A. Needless to say, these figures are only very rough approximations to the ratings you may have actually made.

Beginning on the next page are three additional stories written by subjects drawn from the same population of subjects that you will be rating for the present study. I would like you to compare these stories for practice before you begin the other ratings and then discuss with me your impressions and ask me any questions you may have. Three practice rating sheets have also been provided.

A few cautionary comments before you begin the final comparisons: (a) Care must be taken to accumulate your impressions over <u>all four stories</u> for any subject. For example, it is quite possible that one of the four stories reflects the need for achievement to a moderate or considerable degree while the remaining three seem to reflect very little or vice versa. In the former case, the accumulation <u>across</u> stories must still indicate a relatively small amount of need for achievement for that subject. (b) Try to eliminate from your ratings the influence of any personal bias you may have concerning such things as the person's writing style, grammatical abilities, achievement goals (e.g. to become a champion garbage collector), etc. (c) Any evidence in the stories which you feel may reflect clinical pathology or eccentricity (like perhaps the last story of Subject A from the <u>example</u> stories) is, of course, considered totally irrelevant for present purposes. (d) Try to avoid ratings of <u>complete similarity</u> (A has exactly as much as B) or <u>complete dissimilarity</u> (A has zero need achievement compared to B). (e) Each rating must be made independently of the others; hence, <u>do not refer back to previous ratings for assistance</u>.

Subject A Practice

Story 1

"The two men are experimenting with the development of a new metal which will revolutionize the auto industry.

"Time after time the die has broken and no further advancement can be seen for this new industry. Now after months of frustration it looks as if ... .

"Will this experiment work? How will this material hold up? The other fellow waits in excruciating patience.

"Failure? No, but now the results seem to be better maybe? Just maybe the next time? See the next chapter. Will they?"

Story 2

"The old man turned to the brilliant but erratic scientist. What shall I tell him? What can I do to persuade him?

"The young man had a brilliant career in designing a new program. But had failed.

"I shall succeed said the young man to himself. But this older generation seems to have forgotten how we operate. Maybe I can ... .

"The young generation eventually knuckles under for a period of time waiting. Now the time has come and another generation waits patiently."

Story 3

"Jim wonders, 'Why the math? on a beautiful day? Wonder how I can get finished with this and spend some time thinking about ... .

"It was a warm hot day in June the class had been (?) interminably? I know this stuff he thought? Maybe I can ... .

"Ways of passing time under this dull (?) man passed through Jim's mind? Finally he hit upon a brilliant solution?

"The bell rang waking up a tired boy who sullenly walked out the class? ... Another day has ... ."

Story 4

"The young man was torn between two destinies. To be a soldier or a brilliant surgeon.

"That afternoon they had discussed professions and a chain of thought had been running thru the young mans mind.

"I want to serve my fellow man. How best can I do it. The armed forces with its dullness and sometimes exciting times or the medical profession with its equally ... .

"Time will pass and the moments of history will make the decision. If its war time Jim will have a chance possibly to do the two things he thought of as a young boy. If he survives ... ."

Subject B Practice

Story 1

"Two men, scientists, are about to conduct an experiment, around 1820, on human physiology, and for that reason they are making barbels.

"In the past little has been understood about why people become tired from doing work; these men are attempting to determine why this is so by conducting their experiment.

"They are in the process of preparing their apparatus, so most of the thinking or preparatory aspect has already been completed.

"People of various hereditary backgrounds and of different physical constitutions will be selected (men at this time of history)."

Story 2

"The two men are father and son. The father is trying to persuade -- logically, calmly -- his son to follow him through in the business establishment he has created.

"The son, having become interested in the field of art, does not really wish to become an executive -- but he respects his father and has a difficult decision to make.

"The son wishes not to injure the father's feelings; yet he must make his point.

"The son will select his own field of endeavor; and his father, disappointed, nevertheless wants his son to do as he wishes."

Story 3

"The individual is a junior high school boy in Tennessee. He is staying after school having his teacher explain an arithmetic problem.

"In the past the boy has shown somewhat of an unwillingness to be subjected to the kind of subjects he has had.

"Perhaps the boy realizes that he should work harder at his subject just to please everyone, bu, really, he (?) rather go fishing.

"The boy will continue to go to school, do a mediocre job (which is better than none), and become a farmer in Tennessee."

Story 4

"David McLeash is in the museum in Chicago looking at the exhibits. Behind him is an historical exhibit on medicine. He is looking at the passers-by.

"The McLeash family, from Milwaukee, is visiting relatives in Chicago, and since 10-yr. old David has never been there, they've gone to the museum.

"David is, as a whole, enjoying the vacation, and some aspects of the museum are fascinating to him. But to others he is also indifferent, after all he's 10 yrs. old.

"On the way out of the museum they will become caught in a traffic jam, getting home late for supper."

"(How's that?)"

Subject C Practice

Story 1

"An older man is watching a younger man operate a lathe. The period is in the late 1800's or early 1900's -- clothes, shop and equipment tell me this.

"The two men have had a mild argument: The older man does not believe the younger man can do the job (of turning something or other) well.

"Old man thinks younger can't do the work and is standing by to assert his own superiority. Younger is determined to do it.

"Younger man will do the job and older man will say -- part with regret, part with delight, 'Couldn't have done better myself'."

Story 2

"Grandfather -- polished, urbane, successful, well thought of, is counselling his grandson who is headstrong and not easily awayed.

"Grandson wants to leave college (which grandfather is paying for) to marry a girl who is going to have a child by him.

"Grandfather, worldly-wise, says, 'Don't be ridiculous. She's a tramp. Don't ruin your whole goddam life. Someone of her own social level will come along.'

"Grandson will take grandfather's advice, transfer to another college and become 'successful.' He'll always have guilt feelings about that girl."

Story 3

"The boy is a young, sensitive high school student who dreams of beautiful words, sounds, people.

"He is not very good at school -- not because he is stupid but because his mind is always elsewhere.

"At the moment he is thinking of how the plum trees smell in the spring in his grandmother's back yard. Memory of the scent recalls a specific time he was there and suddenly he can see, feel, hear everything that happened then.

"Teacher will say, 'And Joe, what do you think of this use of the verb?'"

Story 4

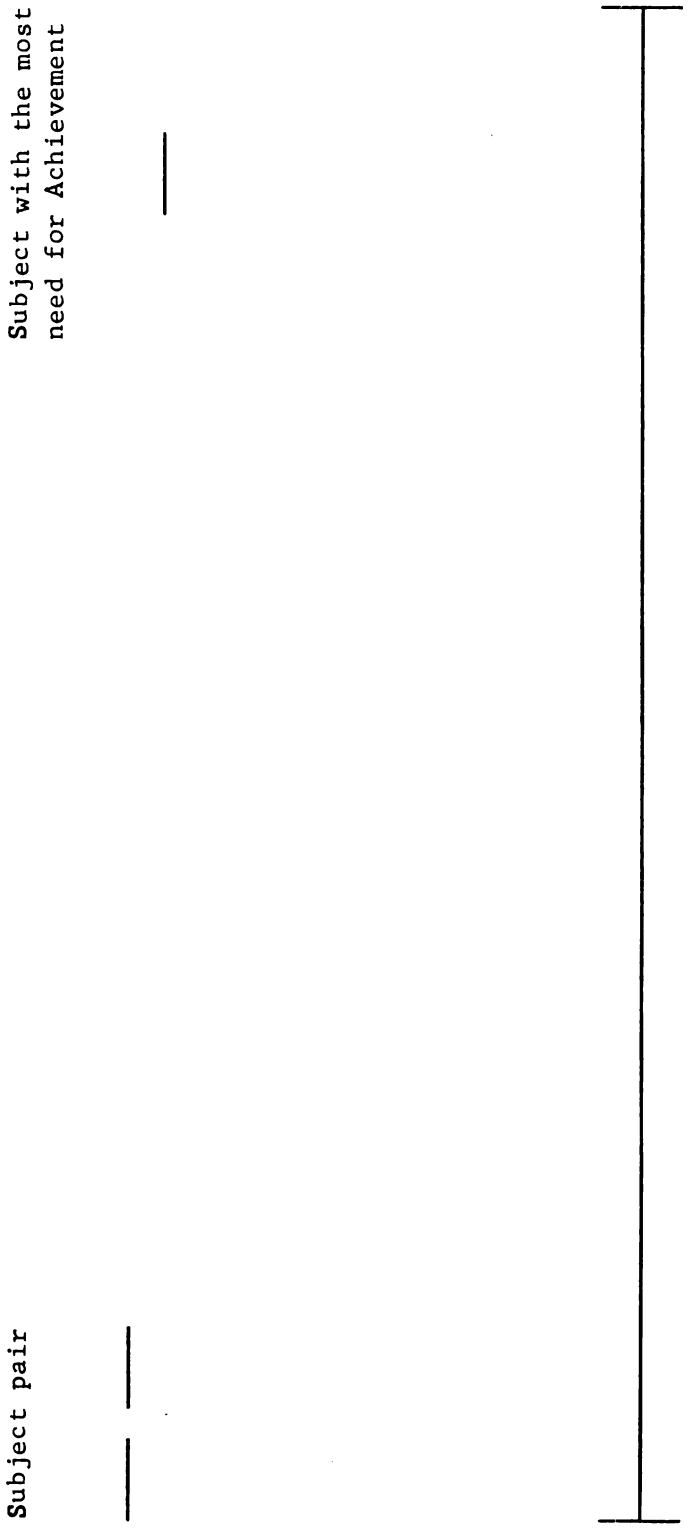
"This is a poster put out by A.M.A. The boy in the foreground is supposed to be thinking, 'How I'd like to be a great surgeon.'

"The boy is from an underpriveleged family, but A.M.A.'s saying, 'You, too, can be a great surgeon. Anyone can succeed in America.'

"The only one doing any thinking is the P.R. man of the A.M.A. who has romanticized surgery (evidently of the 1890's) as ... .

"Nothing except that A.M.A. will go on its reactionary way fighting Medicare, keeping supply of doctors low and prices of med. services high."

Sample Rating Sheet



ORDER FOR COMPARING THE EIGHT PROTOCOL SETS

(1)	A-B	(15)	A-D
(2)	H-D	(16)	C-E
(3)	G-E	(17)	B-F
(4)	F-A	(18)	H-A
(5)	C-B	(19)	E-D
(6)	E-H	(20)	F-C
(7)	F-G	(21)	G-B
(8)	A-C	(22)	A-E
(9)	B-D	(23)	D-F
(10)	H-F	(24)	C-G
(11)	G-A	(25)	B-H
(12)	D-C	(26)	E-F
(13)	E-B	(27)	D-G
(14)	G-H	(28)	C-H