

Utah State University

DigitalCommons@USU

All Graduate Theses and Dissertations


Graduate Studies

5-1962

A Comparative Analysis of the Effects of Heterogeneous and Homogeneous Elementary School Grouping Practices on Selected Personality Variables

Clarence D. Swaner
Utah State University

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>

 Part of the [Child Psychology Commons](#), [Experimental Analysis of Behavior Commons](#), and the [School Psychology Commons](#)

Recommended Citation

Swaner, Clarence D., "A Comparative Analysis of the Effects of Heterogeneous and Homogeneous Elementary School Grouping Practices on Selected Personality Variables" (1962). *All Graduate Theses and Dissertations*. 5569.

<https://digitalcommons.usu.edu/etd/5569>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.



A COMPARATIVE ANALYSIS OF THE EFFECTS OF HETEROGENEOUS AND
HOMOGENEOUS ELEMENTARY SCHOOL GROUPING PRACTICES
ON SELECTED PERSONALITY VARIABLES

by

Clarence D. Swaner

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF EDUCATION

Approved:

UTAH STATE UNIVERSITY
Logan, Utah

1962

ACKNOWLEDGMENTS

The investigator would like to acknowledge the assistance of the members of his committee: Dr. Arden N. Frandsen; Dr. Heber C. Sharp; Dr. David R. Stone; Dr. Terrance E. Hatch; and, particularly, Dr. Walter R. Borg—whose patience, help, and constructive criticisms were greatly appreciated.

The investigator also wishes to express his appreciation to his wife and children who so patiently and constructively supported this effort.

C. D. Swaner

TABLE OF CONTENTS

INTRODUCTION	1
Statement of the problem	1
Hypotheses	3
REVIEW OF THE LITERATURE	5
Ability grouping	5
Summary and conclusions	10
Thematic-Apperception-type tests	10
Methods of interpretation	11
Summary and conclusions	22
Anxiety and methods of measurement	24
The Children's Manifest Anxiety Scale	27
The Test Anxiety Scale for Children	37
DEVELOPMENT OF THE SCORING METHOD FOR TAT-TYPE STORIES	46
Some related scoring methods	47
Development of the scoring method	50
First phase	51
Second phase	58
Third phase	61
Suggestions concerning the use of this scoring method	63
METHODS AND PROCEDURES	66
Method of sample selection	66
Sample composition	69
Development of the testing procedure	70
Methods of scoring	74
Method of statistical analysis	75

RESULTS AND DISCUSSION	79
Scoring of TAT-type stories	79
Findings	79
District comparisons	80
Findings	81
Interpretation	83
Additional district findings	90
Ability level comparisons	92
Anxiety, aggression, and depression	92
Manifest Anxiety as a measure of drive level	96
Text Anxiety	98
Additional level findings	100
Sex differences	103
Aggression	104
Depression	106
Manifest Anxiety scores	107
Test Anxiety	109
(L) Lie scale	111
SUMMARY, FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS	114
Summary	114
Development of scoring method	114
Methods and procedures	114
Findings	115
Scoring of TAT-type stories	115
District comparisons	115
Ability level comparisons	117
Sex differences	118
Conclusions	120
Recommendations	122
BIBLIOGRAPHY	123
APPENDIX	130

LIST OF TABLES

1.	Inter-rater reliability coefficients, Edith Bowen School sample (Spearman-Brown Correction applied)	58
2.	Second phase accumulative inter-rater reliability coefficients (Spearman-Brown Correction applied)	60
3.	Phase three inter-rater reliability coefficients (Spearman Brown Correction applied)	63
4.	Sequential Test of Educational Progress (STEP) cut-off scores used in the selection of the sample	69
5.	Sample composition	69
6.	Sample composition for combined N's across the variable of inferiority	78
7.	Sample composition for the composite (L) scale N's	78
8.	Final corrected inter-rater reliability coefficients	80
9.	Significant district differences	82
10.	Significant ability level differences	94
11.	Significant sex differences	105
12.	Non-significant mean CMAS score sex differences for developmental pupils in both districts	108
13.	Non-significant mean TASC and CMAS score sex differences for developmental pupils in both districts	110

INTRODUCTION

The relative advantages or disadvantages associated with the practice of grouping school children according to ability level has long been the subject of considerable discussion among educators. The general tendency has been for educators to divide into two camps: Those who expound the various merits of homogeneous grouping, and those who favor and support heterogeneous grouping practices. Less common, but acknowledged, are those educators who favor a combination of the two methods. Research dealing with the relative value of either method has usually been concerned with only one aspect of the problem--that of academic achievement. However, the efforts of several investigators (Luchins and Luchins, 1948; Mathias, 1959; Tonsor, 1953) have suggested that different grouping practices may have differential effects upon the social adjustment and personality development of the pupil as well. As the total development of the child is of primary importance in the educational process, an investigation of the differential effects inherent in either of these methods of grouping would seem to be of value.

Statement of the Problem

The primary purpose of this study was to investigate the differential effects of heterogeneous versus homogeneous grouping practices on the personality structure of sixth-grade public-school children. The personality variables of anxiety, aggression, depression, and

feelings of inferiority on a sample of pupils drawn from schools utilizing each respective grouping practice were selected for investigation.

A second purpose of this research was to test the reliability of a partially objectified, inter-rater type, rating scale method of scoring Thematic-Apperception-Test-type stories. The developed rating scale was used to evaluate the personality variables of aggression, depression, and feelings of inferiority. Anxiety was evaluated by two objective-type test questionnaires.

The research design allowed for the investigation of the differential effects of three conditions of grouping practice (district, level, and sex) on the measurement of six selected personality variables. By holding two of the conditions constant and manipulating the third, it was possible to test for significant differences across each of the three grouping conditions. Allowing for the manipulation of one variable, the three possible levels of comparison were:

1. District--in which level and sex were held constant across districts.
2. Level--in which district and sex were held constant within districts.
3. Sex--in which district and level were held constant within districts.

On the basis of a review of the literature and critical subjective analysis, the following hypotheses were set forth in terms of the three possible levels of comparison.

Hypotheses

1. Thematic-Apperception-Test-type stories can be scored by a five-point rating scale with sufficient accuracy to permit inter- and intra-group comparisons on the personality variables of aggression, depression, and feelings of inferiority.

2. In relation to the selected personality variables (anxiety, aggression, depression, and feelings of inferiority) there will be no significant differences between similar pupils in the two districts, except for the following:

a. Superior and developmental pupils in the heterogeneously-grouped classroom will obtain higher mean scores across the variables of aggression and depression than similar pupils in homogeneously-grouped classrooms as an effect of frustrations imposed by the restrictions of a school situation that is geared to the needs of the average pupil.

b. Homogeneously-grouped superior pupils, as a result of intra-group competition and fears associated with losing social status if not successful, will obtain higher Manifest and Test Anxiety mean scores than heterogeneously-grouped superior pupils.

c. Homogeneously-grouped developmental pupils will, as a result of being classified as "dumb," obtain higher mean scores on the personality variable "feelings of inferiority" than similar heterogeneously-grouped pupils.

3. In relation to the three ability levels (superior, average, and developmental), the following significant inter-level differences will be demonstrated:

a. Superior and developmental pupils in both grouping systems will obtain higher mean scores on the personality variables of anxiety, aggression, and depression than average pupils.

b. Superior pupils in either grouping situation, as a measure of "drive level," will obtain higher Manifest Anxiety test mean scores; and there will be a hierarchy of Manifest Anxiety scores proceeding from superior, through average, to the developmental-- in descending order.

c. Developmental pupils will obtain higher Test Anxiety mean scores than average or superior pupils in both districts.

4. There will be no significant sex differences except the following:

a. Boys at all levels will obtain higher mean scores on the personality variable of aggression than girls.

b. Girls at all levels will obtain higher mean scores on the personality variable of depression than boys.

c. Girls at all levels will obtain higher Manifest Anxiety test scores than boys.

d. Girls at all levels will obtain higher mean Test Anxiety scores than boys.

e. Girls at all levels will obtain higher mean (L) lie scores than boys on the Manifest Anxiety (L) scale.

REVIEW OF THE LITERATURE

The review of the literature will be divided into three sections: ability grouping, Thematic-Apperception-type tests, and anxiety and methods of measurement.

Ability Grouping

Research concerned with the advantages and disadvantages of ability grouping can generally be classified into those studies dealing with achievement differences under the two systems, and those studies dealing with social and personality outcomes. As it is the interest of this study to investigate the personality outcomes, only research dealing with the latter area will be presented.

Many of the research studies done in this area seem to have been poorly structured, experimentally unsound, and to have lacked the necessary controls. In fact, a good many of the published articles are merely opinion articles reflecting the personal biases of the writers. As a consequence, it is difficult to attempt to make specific deductions from the published material. Evaluation of the specific effects of different methods of ability grouping on personality variables in the references cited will of necessity be deductive inferences and at best the bases for hypotheses which might stimulate or have stimulated further research.

An interview study of 190 fourth- fifth- and sixth-grade children by Luchins and Luchins (1948) supported some of the critics of ability grouping. This work indicated that in an ability-grouped program

children were aware of the grouping system, and the authors concluded that a stigma is attached to membership in the lower groups. Children who were in the bright group indicated that they would not like to "mingle or be educated with dumbbells" and that they would rather take the chance of failing in the bright group than be in any way associated with the dull or average groups. The study further indicated that many of the dull-class pupils appeared to feel inferior and ostracized. In brief, homogeneous grouping seemed to help create a kind of caste system in the school. In addition, the study suggested that ability grouping, in this case, tended to emphasize learning for external rewards rather than learning as a result of a wholesome attitude toward education. As this study did not include a control group, there is no basis for assuming that caste systems do not also occur in heterogeneously-grouped classrooms.

Tonsor (1953), in an opinion article, suggested that slow learners resent being placed in slow groups, and that grouping tends to limit "social growth" due to unequal social opportunities. He asserted that in everyday life people of all different abilities have to learn to cooperate and that grouping inhibits this necessary social interaction. The inference was that this will have a negative effect on the personality development of the pupil.

Gowan (1955) took an opposite stand and argued that homogeneous grouping practices were democratic and suggested that the homogeneous classroom situation approximates everyday face-to-face contact in work, family, and interest groups--which he concluded are homogeneous. He indicated that the homogeneous group is actually the only kind of group

within which the student can, at first, learn the democratic process of living. He asserted that forcing the child to work closely with those with whom he cannot identify is equivalent to asking a child to read Chaucer--he simply fails in learning social empathy and social skills.

Bettelheim (1958) argued against ability grouping on the basis of discrimination and asserted that the case of the slow learner parallels the Negro caste system in our country and can be expected to produce the same types of anxieties, frustrations, and problems.

Hunt (1942), in a study dealing with the purposes of ability grouping, stated that the purpose of ability grouping, he felt, is to reduce educational waste and inefficiency and to permit the fullest possible development of the individuals involved. The emotional aspects of his report lay in his feeling that the teacher faces

the dilemma of having to decide whether to force the entire class to wait on the slowest, while the average and superior pupils mark time; whether to confine attention to those able to make greatest progress; or whether to make both the superior and least able groups suffer for the purpose of exposing the hypothetical "average" student in minima. [*Italics mine.*] (Hunt, 1942, p. 595)

The connotation that some group will have to suffer is clearly indicated.

Gallagher (1958), in a study of peer acceptance of highly gifted children in an elementary school setting, found that the peer acceptance of highly gifted children (Stanford Binet I.Q.'s above 150) was significantly greater than that of children of average intelligence. Fifty-two per cent of the gifted groups were in the top quarter of their class in terms of sociometric choice, as measured by a simple sociometric device. In addition, gifted children were chosen by children of all levels of ability, and not merely by the bright children. One inference from this

might be that homogeneously grouping the average and dull students might deprive them of the contacts they desire with the brighter children and possibly lead to dissatisfactions and personality problems. On the other hand, removing these bright children from the group may increase the chances of the average or slow child to gain social recognition.

Goodlad (1954) studied the effects of promotion and non-promotion upon the social and personal adjustment of children and found that non-promoted children were rated by their teachers as more bullying (aggressive) than promoted children. The difference was statistically significant. He also found that promoted children were rated by their teachers as cheating significantly more often than were the non-promoted children. This finding may support the inference made by Luchins and Luchins (1948) which asserted that the attainment of good grades was more important than the method by/for which they were earned. In addition, Goodlad found that promoted children revealed indications of being disturbed over their school progress. To relate Goodlad's findings to the effects of ability grouping, one has to assume the condition of being non-promoted or promoted as equivalent to being placed in a high- or low-ability group. This equivalence has not been demonstrated, but the derived inferences may be similar and of interest.

In support of ability grouping was Marsh's (1955) experiment with a special method of screening pupils before assigning them to particular groups. Marsh found that when pupils were rated by teachers over a number of academic variables; matched on the basis of these ratings; and were given a choice of electives, friends, teachers, and class-time scheduling---classroom discipline problems were reduced to

the lowest point since the inception of the school in 1949. Marsh's inference was that grouping can cause problems, but with sufficient planning and programing it can also reduce classroom problems.

Engle (1938), in a study of the effects of school acceleration upon the personality and social adjustments of high school and university students, obtained results not directly related to ability grouping, but that may have some relevance. He found that some accelerated students believed that they had been socially handicapped by their acceleration, and that their difficulty seemed to result from a feeling of being different from their peers. This study seems to be applicable to ability grouping in that it demonstrates the possibility that acceleration may have particular personality effects on the superior individuals, in that people may expect more from them than they may feel able to produce. Likewise, in a dull group, ability grouping may have the opposite effect, allowing the student to be able to compete more adequately. This is not stated in the article, however.

More specifically, Mathias (1959) reported at the American Psychological Association Convention the use of the Thematic Apperception Test (TAT) and Rorschach Test to measure aggression and depression in high, average, and low I.Q. pupils. He concluded that both bright and slow pupils had significantly more aggressive responses than average pupils. He further concluded that the greater amount of aggression found in the bright and slow pupils was due to frustration in the school situation which is aimed primarily at the average student. In relation to depression, the high I.Q. pupil demonstrated the least amount of depression, while the low I.Q. group demonstrated the greatest amount of

depression. No adequate explanation was given.

Summary and conclusions

Mathias' study was poorly controlled; hence, the results were again only inferences, but fairly representative of the type of studies presented in the literature.

The arguments for and against heterogeneous or homogeneous grouping practices are many and diffuse. At the present time no one seems to be able to safely predict what would be the result if either procedure was applied to any sample of school children; however, the available research does raise some interesting questions which should be answered. This study will be an attempt to further clarify this problem.

Thematic-Apperception-Type Tests

The Thematic Apperception Test (TAT) and the Michigan Picture Test (MPT) are techniques used in the investigation of the dynamics of personality as it manifests itself in interpersonal relations and in the apperception or meaningful interpretation of the environment. They consist of a series of pictures about which testees are asked to tell stories. They thereby, presumably, reveal their personal, individual apperception of purposely ambiguous stimuli. As the assumptions underlying the TAT are essentially the same as those underlying the MPT, and as the TAT has been utilized more extensively in research than any other instrument of this type, research dealing primarily with the TAT will be presented--although in the present study a combination of selected TAT and MPT pictures were used.

The TAT is regarded as a projective test, in that the stories which subjects tell about each of the pictures are considered to be projections--that is, ascriptions of feelings and sentiments, needs or drives--of the individual, which are elicited by the stimulus materials of the pictures. According to the projective hypothesis, the mechanism of projection is utilized by the ego as a defense against unacceptable forces and is in part, at least, unconscious. The central assumption of the TAT method, as well as most of its modifications, is that the subject reveals basic elements of his personality in constructing his stories.

Methods of interpretation

Most of the comprehensive systems of interpretation are based upon two elements of the stories: form and content. Form pertains to those elements of the stories which are relatively independent of content and includes such variables as length of stories; degree of organization; responses to details of the cards; inclusive whole response; preponderance of attention to past, present, or future references; and variability in such factors as these from one story to another. Content variables include such molar units as thema or central elements of plot in each story, or somewhat less comprehensive elements, as the kinds of action attributed to various figures--males, females, old, young--and the outcomes of stories--happy, sad, hostile, affiliative, etc. In addition, most interpreters utilize the concept of symbolic content in reference to some form elements of cards seen as having special symbolic implications, or total story situations reflective of various crucial life crises--Oedipal drama, authority resentment, sexual anxiety, etc.

The original technique used by Murray (1943) and his co-workers depended on an analysis of the stories by the Need-Press Method. Every sentence was analysed as to the needs of the hero and the environmental forces (press) to which he is exposed. Every story was so analysed according to all needs and press, and each need and press received a weighted score. A rank order system of the needs and press was then tabulated, and the hierarchial relationships of the needs to each other was investigated. Since its inception, Murray has expanded his interpretation to include a complicated system of Primary (biological) and Secondary (psycho-genic) needs analysis. The analysis is very comprehensive, but cumbersome and time consuming.

As an outgrowth of the original Needs-Press technique, several other authors have developed techniques for scoring TAT-type picture.

Rapaport's method of interpretation (Abt and Bellak, 1950, p. 193) is an examination of the cliché quality of responses. He believed that the subject's deviation from clichés serves as a base line for orientation. In his "points of view" for scoring, Rapaport suggests two main characteristics: (a) formal characteristics of verbalization, and (b) formal characteristics of story content: tone of narrative, figures of story-identifications, memory representations, strivings, attitudes, and obstacles.

Rotter's approach to interpreting TAT stories (Abt and Bellak, 1950, p. 193) is presented in three steps: (a) autobiographical quality; coherence; predominant mood; handling of sex; endings and their relationship to the story; repetition themes; unusual wordings; attitude towards the world; characteristics of central figures; typical methods of solving

problems; and characters that can be identified with mother, father, and son; (b) frequency of occurrence of an idea, unusualness, determination of identification, determination of clichés, and selecting alternate interpretations; and (c) qualitative suggestions for the analysis of personality trends.

Henry (Abt and Bellak, 1950, p. 194) utilized two main classifications of characteristics: (a) Form characteristics: amount and kind of imaginal production, organizational qualities, acuity of concepts, observations and their integration, language structure, intraception-extracception, and relation of story told to total thought content; and (b) Content characteristics: general tone, positive content, negative content, and dynamic structure of content.

Tomkins (Abt and Bellak, 1950, p. 195), in a systematic attempt at a logically consistent analysis of fantasy, distinguished four major categories: (a) vectors: comprising needs, or the quality of strivings for, against, under, by, away from, of; (b) levels: such as those of wish and day dreams; (c) conditions: that may be either external forces or inner states, such as anxiety or depression; and (d) qualities: such as intensity, contingency, and temporal considerations.

Wyatt (Abt and Bellak, 1950, p. 195) utilized 15 variables for the analysis of TAT stories: story description, stimulus perception, deviation from typical responses, deviation from self, time trend, level of interpretation, tone of story, quality of telling, focal figure, other figures, personal relationships, strivings, avoidances, press, outcomes, and thema.

Bellak (Abt and Bellak, 1950) believed that the strength of the TAT lies in its ability to elicit the content and dynamics of interpersonal relationships and the psychodynamic patterns. As a specific inference to the process of interpretation, he stated:

The TAT pictures are best seen psychologically as a series of social situations and interpersonal relations. Instead of responding to real people in real situations, the subject is responding to people in the pictures, which he imagines as certain social situations. Since he is under less constraint of conventionality or reality, his responses are more likely to depict his inner feelings. By this means we get at the contemporary patterns of his social behavior and may be able to infer the genesis of these patterns. Interpretation is the process of finding a common denominator in the contemporary and genetic behavior patterns of a person. (Abt and Bellak, 1950, p. 196)

Bellak's system of interpretation considers fourteen variables: main theme, main hero, attitudes to superior figures, figures introduced, objects introduced, objects omitted, attribution of blame, significant conflicts, punishment (for crime), attitude of hero, signs of inhibition, outcome, pattern of need gratification, and plot.

The materials cited indicate the complex nature of TAT type stories and their interpretation. Considering the complexities and time involved in interpreting stories by a method such as Murray's, Henry's, or Wyatt's, it is understandable that other authors and investigators have attempted to objectify TAT scoring and interpretation. In the interest of specific research, several authors have attempted to develop objective scoring systems for specific personality and behavioral variables.

Some objective-type TAT scoring methods. Little and Schneidman (1955), in an attempt to test the validity of TAT technique interpretations, subjected the TAT and MAPS (Make a Picture Story) protocols of

a particular subject to 17 competent psychologists, who were required to write an essay giving a personality description of the subject. The reports were then separated into discrete statements about the subject, and 802 nonduplicated items were secured. Through an application of Stephenson's Q-sort technique an attempt was made to determine the types of inferences made from TAT stories and the relation between the type of inferences made and their accuracy. Accuracy was determined by comparing the inferences derived from the TAT and MAPS interpretation with the evaluations of 29 workers having detailed information about the subject. While there may be some question as to the accuracy and reliability of this experimental procedure, these authors indicated that clinicians interpreting from a TAT and MAPS protocol can describe a person as accurately as can clinicians working from the elaborate data of a clinical folder.

In another study, Elliott (1955) utilized the TAT in a clinical study of Canadian soldiers who were hospitalized in a base psychiatric hospital for a period of at least eight months. Elliott described the TAT as giving an "X-ray" of the personality, which gave the clinician a look at the emotional life of his patient. In further study, Elliott compared the TAT impressions with interview notes and used the TAT impression pragmatically in the psychotherapeutic situation. In both uses he found a high degree of validity in that the TAT impressions were demonstrated in the clinical situation. The study lacks adequate experimental controls, but suggests concurrent validity.

It is difficult to demonstrate the validity of the TAT in any other than clinical inference situations, as this is its primary use;

however, the extent to which the TAT is used, and the extent to which so many competent people utilize the results derived from its use, suggests the validity of its function.

Eron, Terry, and Callahan (1950), in a study of the normative values of TAT cards, found that specific TAT cards could elicit much more depression than others. Investigation into other personality variables, such as aggression, anxiety, etc., indicated similar tendencies; and a normative scale for certain personality variables that could be elicited from a particular card was established.

Demonstrating that specific TAT-type stimulus variables could elicit stories reflecting specific response sets has seemingly inspired many researchers to utilize the TAT approach in the investigation of specific personality and behavior variables.

In determining the reliability of derived TAT objective-type scoring systems, several authors have selected a personality variable, constructed a list of terms and indices which they felt described the variable, and then rated obtained TAT stories according to the derived criteria. Reliability estimates have usually been in terms of inter-rater reliability coefficients or the test-retest reliabilities of a single or group of raters.

Harrison and Rotter (1945) studied the problem of reliability of the TAT when the same thematic stories were scored by two or more examiners. A sample of 70 officer candidates were administered the TAT by the group method in a semi-darkened room by means of a slide projector. The subjects were shown the pictures for thirty seconds, after which the room was lighted and the subjects were given seven and one-half

minutes in which to write a story. Stories were rated on the variables of emotional maturity and stability by two raters, on a blind analysis basis, across both a three- and a five-point rating scale. The inter-rater reliability coefficient for ratings on the three-point scale was .72, and on the five-point scale, .77. The findings suggested that it is possible to secure good interpretation and intra-personal reliability ratings on the emotional stability of officer candidates between different examiners from a limited number of TAT stories.

Two studies by Fine (1955a and 1955b) describe a scoring system designed to score feelings, outcomes, and interpersonal relationships from the manifest content elicited from TAT stories. In the investigation of feelings, his procedure was to establish 26 response categories, such as: affection, anger, anxiety, compulsion, inferiority, etc., which he felt depicted most of the themes reflected in TAT stories. He then described each response category in terms of descriptive adjectives which were furnished to judges (raters) for use in rating the stories. "Outcome" was described in terms of favorable, unfavorable, or intermediate. "Interpersonal relationship" was rated in terms of interpersonal interaction in which the content of the story indicated a tendency to move towards or away from interpersonal relationships. When the stories were rated by five judges, on the basis of the above-listed criteria, the individual inter-rater reliability coefficients were .79, .78, .70, .68, .64. The judges were given a second trial at rating the same stories, and the resulting reliabilities were .91, .81, .89, .88, .85, respectively. Fine describes his results as positive.

In relation to TAT scoring systems in general, Fine stated:
"The term scoring system for the TAT is a misnomer, what we really have in any 'system' is a shorthand method of extracting meaningful data."
(Fine, 1955b, p. 308)

Dana (1956) studied the TAT records of a group of 150 normal, neurotic, and psychotic females--all of whom had completed one year of college. The variables studied were: perceptual organization, the ability to follow directions; perceptual range, the normality of the response; and perceptual personalization, the rarity of the response. The stories were rated by the test administrator and a clerk, and the reliability was reported in terms of percentage agreement. The percentage of agreement was 94 per cent, 91 per cent, and 76 per cent, respectively. The 76 per cent agreement on the variable of perceptual personalization was thought to be a product of the subjectivity required to make a judgment on that variable. Results of the study indicated that the scoring system was adequate enough to significantly discriminate between all three groups on all three scoring categories.

Auld, Eron, and Laffal (1955) applied Guttman's (1944) scaling method to the Navy version of the TAT (a ten-page booklet with a TAT picture printed at the top of each page). Subjects for the experiment were 100 sailors who were engaged in "Operation Hide-out." (The men spent 30 days in a submarine in an effort to determine the psychological effects produced by different concentrations of carbon monoxide gas.) The obtained stories were rated across the three variables of aggression, sexuality, and dependence. The method of story analysis assumed that a need or drive was shown whenever the subject described an action or

feeling of the hero. The rating procedure consisted of the rater subjectively identifying clear-cut needs or drives and counting the number of times in which the need or drive appeared. Reliability was determined by the split-half method in which the two raters scored 50 of the protocols at one time and 50 at another time. The percentage of agreement between scores ranged from 81 to 97 per cent. A Tau coefficient between raters was computed, and the values ranged from .52 to .92, with a median value of .76. The results indicated that no single card elicited enough responses to utilize Guttman's scaling method across any of the variables. In relation to the utilization of Guttman's scaling method, the results indicated that in order to utilize Guttman's method of scaling, TAT cards must be: (a) selected pictures that would elicit stories of the desired theme, and (b) the pictures selected must tap a unidimensional underlying motivation.

Guttman's (1944) method of scaling demands that the rated need or drive be present in sufficient quantity to be identified as the major theme elicited by the picture. The primary weakness of his method of scaling is that it demands that the stimuli be so structured that it elicits only what the experimenter desires. According to TAT rationale, a given need and/or drive (if present) should be projected into any story, and all that is required is a method refined enough to identify it. In the past, this method has been clinical judgment, objective observation, and subjective inference. Therefore, scales which identify need/drive variables by objective criteria only are open to serious question.

Kagan (1956), Lesser (1958a), Fridman (1957), et al., have done similar studies to those cited above, and the results have all been in essential agreement. All indicate that specific personality variables can be identified by applying appropriate scaling and rating criteria to the interpretation of the TAT-type stories.

Eron, Terry, and Callahan (1950), in a comment about the use of rating scales in interpreting TAT stories, stated:

The use of rating scales affords an opportunity to combine qualitative clinical judgments and a more or less rigorous quantification. It is possible for the ratings to be as subjective and holistic as necessary, but it is essential that the criteria be verbalized (given to the raters) so that the method can be communicated and subsequent raters can utilize the scales with comparable results. (Eron, Terry, and Callahan, 1950, p. 474)

The group method of TAT administration and the effects of length of story. Traditionally, the TAT has been administered as an individual test by a clinician or therapist. Murray (1943) felt that the test should be administered in two sessions, at least 24 hours apart.

Bellak (Abt and Bellak, 1950) felt that the test can be given at any time and in any sequence or amount, and cited an example in which the test pictures were sent home with the subject and the stories were interpreted later.

Auld, Eron, and Laffal (1955) demonstrated the use of a group booklet form of the test, in which significant results were obtained.

Atkinson (1958) and McClelland (1955) have projected TAT pictures to groups of subjects in the investigation of achievement motivation and have achieved highly successful results.

Johns (1960) tested sixth-grade public school children with the TAT by projecting the pictures to groups of students and obtained

written stories which could be scored for achievement motivation. In relation to group TAT administration, Johns felt that sixth-grade students could produce as many as ten written stories to TAT pictures presented to a group by means of an opaque projector.

In a review of the literature, the author found no study which indicated that the TAT could not be profitably administered by the group method. In fact, in current research this appeared to be the preferred method of administration. The popularity of the method is probably due to the economy of time and the larger number of subjects that can be tested.

Most studies indicated that the group method of administering TAT pictures elicits shorter stories, unless a specific number of words are requested. The major criticism to the shorter stories is that they presumably offer fewer cues to the scorer because of the additional effort required to write a story, the restricted number of ideas that can be expressed within a specified time interval, and the absence of opportunity for the clinician to observe and note specific subject behavior during the administration of a test.

Johns' (1960) study with sixth-grade children suggested that written stories were more direct and to the point; and in comparing the number of typed lines written for the ten stories by the pupils in the study, the stories usually did not vary in length more than half a line.

Dana (1955), in a study of the relationship between clinical diagnosis and objective TAT scoring, found that the length of the protocol made no significant difference in the differentiation between

normal, neurotic, and psychotic groups. The inference was that if the variable is present it will be manifest and can be scored if the scoring instrument is accurate and appropriate.

Atkinson (1958), McClelland (1955), et al., all utilized the group method of TAT administration. All reported shorter stories, but all had been able to effectively score these TAT stories for specific variables. It would seem that there can be no real question as to the validity and reliability of the shorter stories, but a question might be directed to the accuracy and appropriateness of the scoring system utilized in the evaluation of the stories.

Summary and conclusions

It is apparent that since its inception by Murray in 1935, the TAT has been subjected to a large number of modifications and alterations in its use as a tool for the study of personality variables. Changes in the method of administration and scoring appear to be the most frequent modifications. In the investigation of specific personality or behavioral variables, these modifications seemingly have not affected the reliability and validity of the material elicited by the TAT stimuli. Of course, as the variables being investigated become more specific and stimulus materials are made more selective, total personality descriptions become more tenuous; however, in the investigation of specific personality variables, the reliability and validity of the findings appear to be quite adequate. The very quality of selectivity which has been demonstrated repeatedly in the literature makes the TAT a useful, realistic, and reliable tool in the investigation of specific personality characteristics.

Eron's (1950) attempts to establish normative stimulus values for TAT cards was a prime stimulus to the investigation of specific personality attributes, since the focus of his study was to demonstrate that certain TAT stimulus pictures were more apt to elicit certain themes than others. Despite the value of his contribution, Eron's inferences have perhaps stimulated too much objectivity in the investigation of personality variables by means of the TAT. Atkinson's (1958) work is a prime example of excessive objectivity. In a purely objective system of scoring, only manifest content is considered scorable; hence in Atkinson's system a content area is scored only if it is stated explicitly. On a purely theoretical basis this is against the basic rationale underlying the TAT technique, which was founded on a subjective, clinical level of interpretation. Since the "new" methods of scoring and interpretation seem to be deriving significantly reliable results, they should, for the present, be recognized and accepted.

However, it is this investigator's opinion that a great deal is lost when strictly objective methods of scoring are applied to the interpretation of TAT stories. This study will attempt to utilize a method of scoring which is objective enough to demonstrate adequate inter-rater reliability and fulfill Eron's (1950) criteria of being "verbalized so that the method can be communicated and subsequent raters can utilize the scales with comparable results," and still retains the clinical subjectivity inherent in the basic TAT scoring rationale.

Anxiety and Methods of Measurement

One of the personality variables to be investigated in this study is anxiety. The term "anxiety" has been widely used with a rather large variety of connotative and denotative meanings. In what appears to be its most general meaning and most common usage, "anxiety" refers to an affective state of the individual--characterized by fearfulness, apprehension, or dread. Following the original suggestion of Freud, many workers have found it both theoretically and practically useful to distinguish between a generalized state of apprehensiveness, in which the individual's anxiety is not restricted to any particular object or class of objects, and fearfulness or apprehension attached to particular objects or situations. Although the custom is not universal, many research workers are tending increasingly to designate the specific types of anxiety which their instruments presumably measure (test anxiety, audience anxiety, sex anxiety, etc.).

In addition to determining the specific type of anxiety which an instrument will presumably measure, it is also of value to define the particular behavioral function that the anxiety performs. Mussen (1960, p. 793) suggests that anxiety must be evaluated both as a motive and also as a response tendency. Response tendency is described as the dispositional quality a subject brings to all situations; and motive, as the pattern or response utilized by the individual that serves the function of reducing, eliminating, or avoiding anxiety. In any appraisal of anxiety in children, one might then ask two basic questions about the anxiety manifest by a child: (a) How anxious and fearful is the

child? and (b) To what extent and in what way does his anxiety serve a motivating function in his behavior?

The two tests of anxiety used in this study, (a) The Children's Manifest Anxiety Scale and (b) The Test Anxiety Scale, are measures of self-reported anxiety and measures of response tendency (situational anxiety). The tests are objectively scored and have the advantage of being group administered.

One might ask the question, "Why wasn't anxiety derived from the TAT protocols in this study, as were the other personality variables?" The response might be that anxiety is an all-pervasive construct that is a part of or accompanies every emotional condition and is therefore very difficult to identify and quantify. Mussen (1960) points out that the global projective techniques, such as the TAT, involve a high degree of subjectivity and aren't as yet well validated as quantitative measuring devices. The two principal advantages of the use of projective techniques for the measurement of anxiety are: (a) The projective methods, because of their indirect approach, may permit the child to express various kinds of anxiety manifestations that he might not reveal directly when questioned or observed; and (b) the projective techniques tend to focus attention upon anxiety as a motive. The primary objection to the use of projective methods as measures of anxiety is their lack of refinement.

Wirt and Broen (1956) studied the relationship between anxiety as measured by the Children's Manifest Anxiety Scale (CMAS) and the clinical concept of anxiety. In a study of 505 fourth-, fifth-, and sixth-grade Minneapolis Public School children, 34 of whom had been referred to the

district's Child Study Department, the authors hypothesized that there would be a positive relationship between anxiety as rated by psychologists after diagnostic study and as rated by teachers and scores on the CMAS. Results of the study indicated that the CMAS was valuable as a predictive instrument, but that it measured something different from what is denoted by the clinical concept of anxiety; and that concepts such as anxiety need much more careful denotation than they have as yet been given.

Goodstein (1954) studied the interrelationship among three measures of anxiety and hostility: The Taylor Manifest Anxiety Scale (TMAS) (the adult form of the Manifest Anxiety Scale from which the CMAS was derived), the Elizier Content Test (a method of scoring Rorschach test protocols), and the Iowa Multiple-Choice Picture Interpretation Test (a multiple choice version of the TAT). The tests were administered to 57 undergraduate college students at the University of Iowa. The results of his study indicated that while all three measurement procedures identified anxiety, all were measuring something different. None of the measurement comparisons were statistically significant.

Mandler, Lindzey, and Crouch (1957) utilized the TAT and the Test Anxiety Scale (TASC) on a mixed group of 77 college males and females to study the relationship between TAT signs of anxiety and scores on the TASC. Their results were congruent with other similar studies (Lindzey and Newberg, 1954) and indicated that the conventional TAT signs of anxiety did not appear to show a high degree of association with an independent measure of anxiety (TASC).

In another study of some tentative TAT signs of anxiety, Lindzey and Newberg (1954) selected 20 of the more common statements cited in the literature as indicative of anxiety, translated each statement into an objective scoring system, and compared each against the scores obtained on a Psychosomatic Inventory by 20 undergraduate Harvard students who had volunteered for the study. Results indicated that three of the statements obtained ratings which were significant at the 5 per cent level of confidence, and 13 of the statement ratings indicated a positive trend toward association. In general, however, the results were described as very discouraging.

It is evident that anxiety as a scorable concept is very elusive. Research suggests that objective measurements, while far from adequate, may be more reliable than those obtained from the projective techniques. At least they are more easily quantified. As the CMAS and TASC test procedures seemed to fit into the present research design, are relatively stable measures of one type of anxiety, and are well accepted as research tools, they are utilized as the measures of anxiety in this study.

The Children's Manifest Anxiety Scale

The CMAS is a widely-used questionnaire measure of children's anxiety. The test was developed by Castaneda, McCandless, and Palermo in 1952 for use primarily with fourth-, fifth-, and sixth-grade children. This 53-item questionnaire represents a children's form of the Taylor Manifest Anxiety Scale which was developed by Taylor in 1951. The TMAS was derived from the Minnesota Multiphasic Personality Inventory, which is a well-known personality test. The questions deal with reported

physiological and psychological concomitants of generalized anxiety, worrisomeness, etc. (such as, "I blush easily," "My hands feel sweaty," "I get angry easily," "I often do things I wish I had never done").

Taylor (1956) makes three statements regarding the construction of the TMAS: (a) It was designed to investigate the role of drive (Hullian Concept) in certain learning situations--the assumption being that anxiety scores are in some way related to drive level; (b) the construction of the test was not aimed at developing a clinically useful test which would diagnose anxiety, but rather was designed solely to select subjects differing in general drive level; and (c) manifest anxiety has been operationally defined only in terms of test scores and will be so employed.

The CMAS, being derived from the TMAS scale, was developed primarily to provide a measure of general drive level, or strength of motivation, in experimental studies on the relationship between drive strength and performance in children. In general the results obtained with the CMAS suggest that it is tapping meaningful anxiety manifestations in children with some validity, and it may therefore be used with profit as a research instrument. Because of the original interest in the scores as measures of drive strength in experimental studies, the criteria against which the test has been validated do not appear to have represented a very broad range of theoretical constructs about anxiety as anxiety; consequently continuing construct validation of the test as a measure of anxiety is highly desirable. Despite the inference that the CMAS anxiety scores are primarily drive level scores, it is interesting to note the overlap between the CMAS, the TASC, and the Audience-Anxiety Scale (AAS)--the latter two being accepted as measures of response (situational)

anxiety. Mussen (1960, p. 797) indicated that intercorrelations among the three tests, based on 421 children in the third to tenth grades, were: AAS .74 with the TASC and .62 with the CMAS; TASC with CMAS, .70. This suggests considerable overlap in the basic constructs underlying the three tests and suggests that while the CMAS was designed to measure drive level, it is also tapping some of the same areas as the TASC—a fact which might help account for some of the similarities seen in the research with each test.

Investigating the concept of manifest anxiety as a reflection of general drive level, Castaneda, Palermo, and McCandless (1956b) compared the performance of 37 fifth-grade children grouped on the basis of high and low CMAS test scores on a complex learning task. The subjects were required to learn several rather complicated series of light button combinations to a criteria of perfect performance. The series were graded in difficulty, thereby measuring the subjects' performance on both simple and complex tasks. Results achieved from the use of the CMAS as a measure of drive level paralleled that which has been obtained on adult groups selected on the basis of high and low TMAS scores. In general there is a tendency for the high anxious subjects to perform more poorly in comparison to low anxious subjects as the difficulty of a task increases. Conversely, if the difficulty of the task can be sufficiently decreased, high anxiety appears to facilitate the learning process.

I. G. Sarason (1957) studied the differential effects of anxiety and two kinds of failure situations on two groups of high and low anxiety college students selected from the extremes of a TMAS score

distribution. Subjects were presented two lists of 17 nonsense syllables by means of a Hull-type Memory Drum, and the measure of learning was the number of correct anticipations. The subjects were randomly assigned into three groups, two of which were given failure reports on their progress at various points during the experiment, and one which constituted a control (neutral) group and received no failure reports. Results indicated no significant difference in learning ability between high and low anxiety subjects of the neutral group, but showed that the performance of the high anxiety group was markedly impaired by the presentation of reports of failure. Further analysis indicated that learning ability was impaired for both the high and low anxiety groups after the presentation of failure reports, but the impairment was much greater in the high anxiety group.

Using the CMAS as a measure of motivation, Palermo, Castaneda, and McCandless (1956) studied the relationship of anxiety in children to performance in learning tasks involving complex discriminative ability. Thirty-six fourth-grade children, selected from the extremes of a CMAS score distribution, were assigned to high and low anxiety groups and required to learn a complex system of colored light discriminations. Results indicated that in complex learning tasks, high anxiety children demonstrated inferior performance when compared to the performance of low anxiety children on the same task.

Taylor (1958), in an attempt to study the effects of anxiety level and psychological stress on verbal learning, subjected two groups of high and low (TMAS) anxiety college students to a neutral and stress (failure report) situation and measured the effect on the learning of

lists of nonsense syllables. The experiment was designed to test the TMAS assumption that obtained scores would be related to the level of emotionality and drive. It was postulated that if TMAS scores were a measure of drive level, subjects with higher anxiety scores would demonstrate greater learning efficiency if the interfering extra task responses to competition and fear of failure were eliminated. Drive theory would suggest that under neutral conditions high anxiety (high drive) groups would perform at a higher level than low anxiety (low drive) groups. Likewise, if the introduction of stress resulted simply in an increase in drive level, and further if high anxiety subjects are more reactive to such stress, these subjects should increase their margin of superiority over the low anxiety groups. If, on the other hand, the major effect of stress is to cause competing extra task responses, the high anxiety group should no longer exhibit a performance superior to the low anxiety group (demonstrated in other research) and may even be inferior to them. The subjects were presented lists consisting of eight pairs of nonsense syllables by means of a Hull Memory Drum. Two groups of high and low TMAS scoring students were divided into two subgroups, one being a stress group and the other a neutral group. Each group was told that the performance on learning tasks was related to intelligence, as a means of stimulating motivation. The neutral group was given each list with no comment other than "Here is another list," while the stress group was repeatedly told that their performance was considerably poorer than most people's. Results indicated that under neutral conditions the high anxiety group exhibited the superiority in performance predicted by drive level theory. Both subgroups, when compared to a control

group, demonstrated a significant decrement in performance when told that their performance had been inadequate. Under the stress situation, high anxiety subjects demonstrated a significantly greater learning decrement than did the low anxiety subjects.

The results of these studies may have inference value to the situation of homogeneous and heterogeneous school grouping practices. It is possible that the grouping practice would represent a situational or psychological stress which might impair learning ability. As an example, a child grouped in a low achiever group might consider this an indication of failure and this reaction might further reduce his learning potential. Research in this area would be of value.

Studies dealing with the relationship between manifest anxiety level, measures of intelligence, and school achievement usually demonstrate negative correlations between the three variables.

McCandless and Castaneda (1956) studied the relationship between anxiety, as defined by the CMAS; academic achievement, as measured by the Iowa Every Pupil Test (IEPT); and intelligence, as measured by the Otis Quick Scoring Mental Ability Test (Form B) on a group of fourth-, fifth-, and sixth-grade school children. Using a multiple correlational technique and partialing out selective variables, the authors reported a negative relationship between CMAS anxiety and both school achievement and intelligence. Results also indicated that as anxiety scores increased, the more complicated school subjects--such as reading, arithmetic, and their composite performance--suffered more interference from anxiety than the more simple skills, such as spelling. In addition, girls seemed to demonstrate more interference than boys, with the one exception of sixth-grade boys in arithmetic. A possible explanation of this may be the

obtained, higher mean anxiety scores for girls and stronger academic motivation. This is in keeping with a previously cited study (Castaneda, McCandless, and Palermo, 1956b) in which high anxiety subjects demonstrated a greater reduction in learning ability than subjects with lower anxiety scores in the same situation.

Matarazzo, Ulett, Guze, and Saslow (1954), studying a group of college students, found no significant relationship between a criterion of grade-point average and TMAS scores. Relating TMAS scores to the timed test situation encountered in the American College Entrance (ACE) test, the authors' data indicated a curvilinear relationship between intelligence scores derived from the ACE and TMAS scores. This suggested that subjects of both extremely high and extremely low TMAS scores tended to earn lower scores on the ACE test. However, when the categories of high and low anxiety were broadened, the data gave a better description of the performance and clearly indicated that as the TMAS anxiety score increased, scores on the ACE test decreased. As the ACE is a timed test, the authors, on the basis of clinical experience, hypothesized that high anxious subjects do not function efficiently under the pressure of time, and that the time factor could have accounted for the negative relationship between ACE and TMAS scores. To test this hypothesis, the authors utilized the Wechsler-Bellevue (Form 1) subtests--comprehension, vocabulary, and similarities (CVS)--as a measure of intelligence and compared these scores to the TMAS scores of the group. The results showed no significant difference between CVS scores and high or low TMAS scores. This suggests that timing (expectations) may be a factor in the negative relationship between ACE and TMAS anxiety scores. The results also demonstrate that

some measures of intelligence are apparently not related to anxiety level as measured by the TMAS. The inference is that the TMAS anxiety is strongly related to situational anxiety and probably involves a personal interpretation by the subject.

In a study testing the reliability of the findings of Matarazzo, et al., Schultz and Calvin (1955) re-did their experiment and found no significant relationship between ACE intelligence and TMAS scores. They attributed the differences to sampling differences, but concluded that no definite relationship between anxiety as measured by the TMAS and I.Q. scores on timed tests had yet been established.

Davids and Eriksen (1958), in another investigation of the relationship between manifest anxiety (TMAS) and intellectual attainment, compared the TMAS scores of 40 male, undergraduate college students with their grade-point averages and the results of a battery of college aptitude tests (not described). The authors failed to obtain any significant differences between TMAS scores and any measure of intellectual performance. They concluded that scores on the TMAS are largely independent of intelligence.

The described research places much doubt around the relationship of manifest anxiety as measured by the TMAS and intelligence. However, the literature clearly points out the relationship between personal involvement and manifest anxiety. It seems clear that whenever a subject places a subjective meaning upon a complex stimulus variable to the effect that the manifest anxiety score is increased, the performance is subsequently impaired.

The CMAS Lie (L) Scale. The CMAS contains an 11-item Lie (L) scale as a measure of response set. The (L) scale is assumed to reflect the tendency for any given subject to falsify his response to the anxiety text items. The (L) score is high for children who deny that rather universal truths are characteristic of them; e.g., "I do not lie," "I am always good," "I never get angry," etc. The initial standardization data presented by Castaneda, McCandless, and Palermo (1956a) indicated that the (L) scale measures something different from manifest anxiety. Attempts to correlate the anxiety scale and the (L) scale for different grade levels and for boys and girls separately, elicited correlations clustering around zero. As with the anxiety scales, there was a general tendency for girls to receive higher scores in comparison to boys.

Most researchers who utilize the CMAS do not utilize the (L) scale, as the items have been demonstrated to have little in common with the direct measurement of manifest anxiety--but more specifically because eliminating subjects who had high (L) scores from a study would eliminate a good part of the total N of many experimental groups. McCandless and Castaneda (1956b) in one study, and Castaneda, Palermo, and McCandless (1956) in another study, stated that eliminating subjects with (L) scores above seven would have reduced the size of their sample N's about 30 per cent.

In the present study, the experimental samples were sufficiently large to allow for the elimination of subjects with (L) scores over seven and still maintain the desired size of the experimental sample. The rationale for doing this was that it seemed desirable to work with subjects who were at least willing to attempt telling the truth and who were motivated enough to take an interest.

Some of the research that deals with the (L) scale is of interest. Trent (1957), studying the relationship of anxiety to popularity and rejection among institutionalized delinquent boys, utilized the CMAS and a sociometric choice questionnaire to study the relationships among 63 delinquent boys in a New York State training school for boys. Results indicated that these institutionalized boys obtained higher CMAS (L) scale scores than the subjects reported in the original standardization study. The (L) score was found to be related to age and I.Q.--the older boys having a tendency towards more falsification, and the lower I.Q. individuals obtaining higher (L) scores than the remainder of the group. From a sociometric point of view, the more anxious boys were inclined to be less popular and more often rejected.

In a previously cited study, McCandless and Castaneda (1956) found high anxiety and high (L) scores related, in that high (L) scores were related to poor academic achievement in much the same manner as the negative relationship between high CMAS scores and school achievement.

In an item analysis of the CMAS, Hafner and Kaplan (1959) analysed the response tendencies of 122 fifth-grade middle-class school children in St. Louis, Missouri. Contradictory to other studies indicating that girls score higher than boys on anxiety and (L) score items, the authors found that individual items generally did not discriminate between sexes. In terms of sex differences, only three of the test items, including the (L) scale items, significantly differentiated between boys and girls. It was interesting to note that the anxiety items answered "Yes" most frequently by the total group of subjects were the same items most frequently answered by the low anxiety group. Despite the inability of

individual test items to differentiate sex differences, repeated research (McCandless and Castaneda, 1956; Castaneda, McCandless, and Palermo, 1956a) indicates that in response to the test as a whole sex differences are demonstrated with girls manifesting higher CMAS and (L) scale scores than boys.

The Test Anxiety Scale for Children

Mandler and Sarason (1952b) developed a 30-item "yes-no" questionnaire (TASC) which is concerned with attitudes toward and experiences in test and test-like situations. Like the Taylor, the TASC is intended to measure anxiety as a personality variable. The hypotheses that have been tested with the TASC focus for the most part upon anxiety as response tendency, since the test questions deal primarily with affective, physiological, and motoric manifestations of anxiety; such as, "Do you worry a lot before you take a test?" "When the teacher says she is going to find out how much you have learned, does your heart begin to beat faster?" "When you are taking a test, does the hand you write with shake a little?" etc.

The TASC is built around three primary hypotheses, and the research reported in the literature are primarily tests of these hypotheses, which are:

- A. The reaction of the test anxious child to test and test-like situations in the classroom reflects his experience in psychologically or interpersonally similar situations in his home both before and after the beginning of formal schooling.
- B. The test anxious reaction has, in addition to its conscious significance, a concurrent unconscious significance which primarily relates to what has been experienced by the child in the family situation.

1. One of the aspects of the matrix of unconscious factors related to the test anxious response is strong hostility towards parents and surrogates whose evaluations of the child's performances elicited in him hostility which could not be satisfactorily expressed. If the hostility was expressed, it was punished; if it was expressed in phantasy, it resulted in conflict with positive feelings toward these figures. It is the sense of this formulation that the instigation of hostility in the child by parental behavior has taken place over a period of years during developmental periods when the child is relatively unable to regulate the strength of response to instigated hostility (a kind of all or nothing mechanism) at the same time that the disparity of strength between parent and child is greatest.
2. Parental handling of the child's hostility may have various effects on the child's attitude towards such hostility, but the most frequent effect is to produce the experience of guilt over his hostility, an effect which is reinforced by the strong positive feelings the child has toward his parents. To the extent that this hostility can be kept unconscious the child avoids the upsetting experience of guilt, i.e., the awareness of how "bad" he is. We assume that in the case of the test anxious child parental handling resulted in inordinately strong hostility and that attempts to defend against its expression were unsuccessful to the degree that it did not avoid the upsetting experience of guilt.
3. The concurrent conscious experience of hostility and guilt makes it likely that the child's attitude toward himself will contain a consciously derogatory flavor, especially if, as in the case of the test anxious child, his hostility and guilt are aroused in situations in which judgment is being passed on his adequacy, i.e., situations in which some assessment is made of the disparity between the child's behavior and parental expectations. What is implied in this formulation is that the test anxious child is one who derogates his own worth and tends to direct aggression toward himself rather than others, a self-attitude which bespeaks the strength of the unconscious hostile tendencies toward others.
4. Another of the unconscious factors increased in strength in the test anxious reaction concerns unconscious phantasies about the consequences of directing strong hostility toward parents. Essentially these are unconscious phantasies of retaliation on the part of the parents--what they will do to the child in response to his hostility--as well as of being in a state of abandonment and helplessness, i.e., a state in which his dependency needs will not be satisfied. It is, in fact, this perceived threat to the fulfillment of his dependency needs which not only serves as a control

against the overt expression of hostility but also motivates behavior which will insure the possibility of satisfaction of his dependency needs. As a result, the dependency of the child on parents for approval, direction, and support becomes a dominant tendency. This in turn would have the effect of inhibiting the child's spontaneity and creativity. Conforming to the expectations of others is, however, in the case of the test anxious child, no stable solution because, it is assumed, his relations with adults are such that his hostility is frequently being aroused in situations where negative judgments about the adequacy of his behavior are being made--and a kind of vicious circle of behavior starts in again. It is important to emphasize that this vicious circle consists not only of the child's repeated efforts to stabilize his conflictful relationship with his parents but also the parents' erroneous perceptions and conceptions of the significances of those efforts. (S. B. Sarason, et al., 1960, pp. 13-22)

- C. The rate of increase in performance of the test anxious child would be significantly less than for the non-test anxious child. This is not to say that test scores of test anxious children would not increase (this might happen in some cases), but that the amount of gain would not be an accurate reflection of potential or capacity. At this point it should be borne in mind that the test anxious response is essentially a covert one, i.e., it is for the most part a complex of responses which is not articulated or clearly discernible by another person. Consequently, it is not likely that the relation between test anxiety and test performance would be recognized so that corrective steps could be taken, assuming that in the classroom situation, for example, one would know what the corrective steps should be. What we are saying here is that in the situation where test anxiety is experienced most frequently (i.e., the classroom), the significant adult tends not to recognize the problem and therefore can be of no help in minimizing the strength and effects of the test anxious response. The passage of time tends not to be therapeutic and the problem may remain at a particular level or become worse. In addition since the test anxious response essentially prevents the child from an objective assessment either of himself or the external situation and tends to result in self-defeating behavior, repeated test experiences would not be expected in themselves to result in a reduction of test anxiety. (S. B. Sarason, et al., 1960, p. 23)

As a summary of what the stated hypotheses and corollaries evidenced in the light of investigation, S. B. Sarason stated:

1. One of the most consistent findings is that the majority of the so-called fears of children are of an imaginary nature, i.e., they are obviously remote from the child's personal experiences.

2. The fears of children are in many respects similar to the fears of parents . . . consistent with our general position that the anxieties of children cannot be understood without relation to the parental variable.
3. It seems clear from these studies that although the frequency of fears remain relatively constant at the different age levels, the content of the fears seems to change systematically as the child gets older.
4. The most consistent findings in all these studies is that the total incidence of fears is higher in girls than in boys. (S. B. Sarason, et al., 1960, p. 40)

Sarason and others have administered the TASC to large numbers of school children from grades one to five, and have diligently explored the relationship between their test anxiety measures and a wide variety of psychologically relevant criteria.

Sarnoff, S. B. Sarason, Lighthall, and Davidson (1958) demonstrated that the TASC has satisfactory reliability by administering the test to 100 randomly-selected school children in grades one, two, and three. The split-half coefficient of reliability was .81. S. B. Sarason et al. (1958) cite data which are essentially the same.

Mussen (1960, p. 794) indicated that the over-all results of the research done with the TASC questionnaire suggests that it is tapping a meaningful anxiety manifestation in elementary school children.

Childs (1954) and Jones (see Eysenck, 1961, p. 502) are in agreement.

In an initial study designed to test the developmental progress of the TASC and check its reliability, Sarason et al. (1958) tested 1,697 mixed students through grades two to five. Besides establishing the reported split-half reliability of .81, the authors also found that (a) test anxiety (TA) scores tend to increase with grade, but that the increase was not steady; (b) high TA scores tend to be negatively

correlated with both I.Q. and achievement; (c) teachers' ratings of anxiety and TA scores are positively correlated; and (d) that children's self-ratings of TA were more valid than teacher TA ratings as grade increased. These results are roughly parallel with those obtained on CMAS research and suggest a similarity between concepts measured by the two tests.

Sarnoff et al. (1956) investigated both the effect of examination immirence upon TA scores and the relationship between TA and pupil performance on the British "Eleven Plus Examinations." They pretested the fourth-grade population of two English schools in June, 1956, and tested the same population when they were fifth-graders in February, 1957, with the TASC. The purpose of the double administration was to obtain a TA measure considerably before and then immediately before the Eleven Plus Examinations. The findings indicated a decrease, rather than the predicted increase, in TA scores. Further research indicated that decrease in TA score was due to "order (position) effect," which suggests that regardless of country, age, or sex, the mean TA scores of all groups--English and American--tend to decline from test to retest. Retesting is assumed to make people defensive with respect to the emotionally-loaded variable measured by the TA scale.

In an effort to investigate cross-cultural variations in TA correlates, Sarnoff et al. (1958) attempted to match 579 American and 533 English children from grades one to five across the variables of test anxiety, general anxiety, lie status, sex, and grade in school. England was selected as the comparison country, primarily because of their Eleven Plus Examinations, which have no counterpart in our country;

and, secondly, because of the language similarity, ease of test data communication, and comparability of certain English school districts with selected American school districts (Connecticut). Because the Eleven Plus Examinations constitute an extreme threat to the average English child, it was hypothesized that (a) English children would demonstrate a generally higher level of TA than their American counterpart; (b) there would be a general rise in the TA as grade level increased (based on the assumption that as examinations increase in number and as academic performance is evaluated with increasing discrimination--which is assumed to be the case as grade level increases--anxiety about performance on tests and in test-like situations will increase); and (c) because both cultures condone freer emotional expression in girls than in boys, girls would have higher TA scores than boys. Results were in keeping with the hypotheses, in that English children demonstrated higher TA scores than American counterparts (significant at the 1 per cent level); girls demonstrated higher TA scores than boys, and TA increased with grade. The study served to demonstrate the construct validity of the TA scales and supported the findings of previous research.

Zweibelson (1956) studied the effects of fear, worry, and anxiety on group test performance. Investigating the relationship of TAT scores to group I.Q. scores and "position effects," he tested five classes of fifth graders with the TASC and compared their scores to their performance on the Otis (Alpha and Beta) and Davis Eells tests. Results indicated that low TA groups made higher mental ability indices than the high TA groups, regardless of the order of test administration and

regardless of the test instrument. In regard to position effect, findings indicated no significant differences between the order of Davis Eells or Otis (Alpha or Beta) test administration. However, there was a tendency for high TA pupils to do better on the test taken second, than first. This finding is in keeping with the findings of Sarnoff et al. (1956) in their test-retest investigation of TA scores preceding the British Eleven Plus Examinations, wherein they demonstrated that test scores on the second administration are always better if other variables are held constant.

Mandler and Sarason (1952a), investigating the role of drive states in a testing situation and the role of anxiety to the effects of subjective success and failure on the anxiety state and test performance, worked with the extremes of a distribution of 101 introductory psychology students ranked on an anxiety continuum established on the basis of their responses to a 67-item anxiety questionnaire. TA level was established three and one-half months prior to the time subjects were asked to participate in the present experiment to avoid the effects of position effect and interaction. No mention was made of the connection between the two situations. In the final experiment, all subjects knew they were taking I.Q. tests, but the TA extreme groups were divided into three subgroups, each of which received either success, failure, or neutral anxiety stimulation while performing on two types of intelligence measurements-- a KOH block design and a modified Wechsler-Bellevue digit symbol test. In support of Zweibelson's (1956) findings, which indicated that TA scores decrease with repeated testing, the authors found that low TA subjects did better on the KOH block design than the high TA group for

the first five trials. However, as the learning process proceeded, the anxiety drive of the high anxiety group tended to improve performance scores. The authors found the variability of the high TA group significantly larger than that of the low TA group, and that an intervening report (success or failure) elicited improved performance for the low TA group, but depressed scores for the high TA group. It appeared that the optimal conditions for a high TA group are those in which no further reference is made to the testing situation, and that the optimal conditions for a low anxiety group are those in which the subjects are given a failure report.

The TASC may have a special relationship to any study concerned with anxiety as a function or result of ability grouping practices, as S.B. Sarason and his collaborators (1960), as the premise for their first basic hypothesis, point out the relationship between children's feeling of anxiety and parental attitudes and influence. It is apparent in their writings that the child brings to the test and/or school situation the basic predisposition for anxiety, which then may be elicited by the school and/or test environment.

It is significant that S. B. Sarason states in another reference:

In our conception, test anxiety is not something that is created in the school, it is something that the child brings, more or less to the school and in that situation can be either made higher or lower or kept at the same level. (S. B. Sarason, 1961, p. 23)

In relation to the present study, this might suggest that children raised in similar cultural and economic environments might, due to the homogeneity of the environmental influences, manifest similar parent-induced anxiety predispositions which might then be brought to the classroom. Assuming these assumptions to be valid, it would appear that

differences in TA scores between the experimental districts used in this study would likely be an effect of the school situation. It is difficult, if not impossible, to assume homogeneity of parent-child relationships simply because the cultural sociometric backgrounds may be similar, but one needs to be aware of this possibility if the true value of the TASC is to be realized.

Conclusions. Considering the results of various studies done with the TASC, it seems justified to conclude that subjects with high test anxiety differ from the subjects with low test anxiety in the habits of responding to anxiety which has been built up through the course of their lives. High anxiety subjects evidently develop habits of responding to anxiety with various responses, internal and external, which are incompatible with efficient pursuit of a complex task; hence these subjects do poorer in a situation which evokes much anxiety than in a situation which does not evoke much anxiety. Low anxiety subjects, on the other hand, evidently lack strong habits of responding to anxiety with task-irrelevant responses, so that anxiety-arousing instructions have as their main effect on these subjects an improvement of performance through increase in drive. In addition, the studies relating to the reliability of the TASC as a useable instrument seem adequate enough to allow its use as a measure of anxiety when anxiety is defined as a "response tendency."

DEVELOPMENT OF THE SCORING METHOD FOR TAT-TYPE STORIES

Several investigators have attempted to develop objective-type rating scales for use with TAT-type stories. Interpretive methods range from a ten-hour psychoanalytic interpretation of a TAT protocol to a simple word count of suggestive adjectives. In dealing with large numbers of TAT protocols, the psychoanalytic method is too cumbersome and time consuming to be practical, and for the serious clinician or researcher the adjective-word-count method seems too superficial and limited to be of much value. Most writers consider the TAT to be a rich source of information concerning the attitudes, feelings, and behavioral adjustment of an individual; and most writers are convinced that the TAT should not be used as a diagnostic instrument, but rather as an impressionistic guide to an individual's feelings and behavioral adjustments.

The present study required the utilization of a large number of subjects (338), all of whom wrote 7-minute stories to seven selected TAT-type pictures which were presented to selected groups of sixth-grade children by means of an opaque projector. The obtained TAT protocols were to be scored for three personality variables: aggression, depression, and feelings of inferiority. The large number of protocols to be scored, plus the desire to study selected personality variables and not total personality structure, influenced the writer to rule out formal analytic-type interpretations and search for a more rapid method of scoring which would be reliable, valid, and retain enough subjectivity to allow some

measure of depth interpretation. A review of the literature concerning some of the established objective-type TAT scoring methods revealed no single scoring procedure which seemed to meet the needs of the study, and the decision was made to develop a scoring method more appropriate to the objectives of this study. To demonstrate that this is not a unique approach, several examples of relative research will be cited.

Some Related Scoring Methods

Eron, Terry, and Callahan (1950) investigated the use of a rating scale in determining the emotional tone of TAT stories. The authors felt that the particular stimulus properties of a particular TAT picture might elicit stories of a certain emotional tone regardless of the clinical classification of the subject. The TAT was administered to 100 college students (50 male and 50 female) according to the standard method as defined by Murray (1943). One thousand TAT stories were rated for emotional tone by three judges on a five-point rating scale that ranged from very sad (-2) to very happy (+2); a rating of zero (0) was considered to be neutral. The inter-rater reliability between the three judges was .80. The results indicated that few of the TAT cards elicited stories which varied significantly in feeling-tone among subjects. This finding suggested that the pictures themselves may be more important than the projections of the subject in determining the actual emotional tone of the story told. The findings suggested that TAT pictures elicited primarily sad stories, but that each picture had its own stimulus value which influenced the emotional tone of the story it elicited. Eron's findings suggested that TAT pictures will, depending

upon the stimulus property of the particular picture, elicit an emotionally-toned response that is descriptive of a particular feeling.

Several investigators employing these findings have attempted to investigate specific personality variables. Eron's study also demonstrated that a rating scale approach could be applied to the identification of emotional tone, and several of the more recent researchers have attempted to develop rating scales designed to investigate particular personality variables.

The measurement of fantasy aggression has been the particular interest of several investigators. Kagan (1956) attempted to measure overt aggression from fantasy responses given to 13 TAT-type pictures by 118 first, second, and third graders (median age 7 years 9 months) from a middle-class, socio-economic environment. Each child was rated by his teacher on two variables: (a) the tendency to start fights, and (b) the tendency to hold in and not express anger overtly. On the basis of the teachers' ratings the children were separated into five groups, ranging from most to least aggressive. The TAT stories were scored for aggression content according to five themes: (a) fighting between boys, (b) destruction of property, (c) stealing, (d) swearing, and (e) physical aggression. Results indicated that boys rated by teachers as most likely to initiate fighting behavior produced significantly (1 per cent level) more fighting themes than those boys rated as extremely non-aggressive. These findings suggest a positive relationship between overt and fantasy aggression; i.e., those boys who demonstrate overt aggression also give aggressive responses to TAT-type pictures designed to elicit aggressive themes.

Hokanson and Gordon (1958) studied the expression and inhibition of hostility in imaginative and overt behavior as revealed by responses to TAT cards by 40 male college students. The study had many methodological weaknesses which limited the value of the findings, but is cited here in relation to the utilization of an 8-point rating scale in the measurement of expressed and inhibited hostility. The 8-point scale ran from a zero score (no expression of hostility) to a score of seven (intense expression of hostility). The stories were ranked by the authors and two psychology graduate students with an inter-rater reliability coefficient of .79 to .80. The authors concluded that the scale was reasonably reliable.

Fridman (1958), in a study entitled "Objectifying the Subjective," used a rating scale based on principles similar to the method utilized in the present study. Fridman was concerned with the problem of reliability when one attempted to deviate from the subjective (holistic) method of interpreting TAT stories in favor of the objective (atomistic) method of scoring. He felt that objective scoring sacrificed validity for reliability; and subjective scoring, reliability for validity. Quoting Eron, Terry, and Callahan (1950), Fridman attempted to justify the use of rating scales in the scoring of TAT-type stories:

The use of rating scales affords an opportunity to combine qualitative, clinical judgments and a more or less rigorous quantification. It is possible for the ratings to be as subjective and holistic as necessary but it is essential that the criteria be verbalized so that the method can be communicated and subsequent raters can utilize the scales with comparable results. (Eron, Terry, and Callahan, 1950, p. 474)

The combination of qualitative clinical judgment and rigorous quantification suggests that rating scales could be valuable tools in

research using TAT-type materials. It has been the practice of several investigators (Guttman, Atkinson, etc.) to rate each TAT story for a single variable, so as not to distract from the reliability of the rating procedure. Fridman (1958) attempted to maintain the holistic approach to the understanding of a personality by rating a single story according to several variables simultaneously. His method of scoring involved the use of 80 prepared statements which described characteristics of the "hero." The design required the use of a single 80-item rating scale in the evaluation of several variables. The stories were rated for outstanding qualities of the hero's behavior by five judges, with a resulting combined inter-rater coefficient of .74. Fridman's procedure differed from that used in the present study in that the present study used a separate list of characteristic "helps" for each personality variable investigated.

Other investigators (Mussen and Naylor, 1954; Kagan, 1956; Lesser, 1958a) have employed various methods of objectively scoring TAT stories, and all seem to agree that TAT-type stories can, with various degrees of reliability, be scored by the rating scale method.

Development of the Scoring Method

The scoring method developed for use in this study was begun by Mrs. Luna Brite (graduate student at Utah State University) and later expanded through the combined efforts of Mrs. Brite and the investigator. The decision to use a rating scale method of evaluating TAT-type stories was based on the following conditions and assumptions: (a) there were a large number of stories to be scored, and the time factor was important;

(b) research had shown the rating scale method to be valuable in the investigation of specific personality variables; and (c) there was a challenge involved. It would have been expeditious to have utilized one of the existing scoring methods, but none of the methods met the needs of the study in the investigator's approach to the problem. A method was desired that would allow quantification of qualitative data, would be rapid in utilization, would allow clinical subjectivity, and would be both reliable and valid. In short, what was needed was a partially objectified, subjective, inter-rater type rating scale that would (a) take advantage of cumulative data elicited from the stories; (b) allow for the simultaneous scoring of more than one personality variable; (c) not be dependent upon concrete response, word counts, or other structured approaches; (d) allow the use of clinical insight, conclusions, and interpretation; (e) elicit no negative scores; and (d) allow all impressions to be additive.

The development and utilization of the method involved three phases.

First phase

In the first phase, Mrs. Brite, with the aid of several individuals who were familiar with the TAT method of personality investigation, selected nine cards from the TAT and Michigan Picture Tests which they felt would elicit stories depicting aggression, depression, and feelings of inferiority from fifth- and sixth-grade school pupils. The cards selected were:

<u>Card</u>	<u>Test</u>	<u>Card Number</u>
1.	TAT	1
2.	Michigan Picture	6
3.	TAT	3 BM
4.	Michigan Picture	3
5.	Boys Girls	10 B 10 G
6.	Boys Girls	8 B 8 G
7.	Michigan Picture	9
8.	Blank card, either test	-
9.	Michigan Picture	2

The test (9 cards) was then administered by the usual method to the 26 fifth-grade pupils at the Edith Bowen School (Utah State University Campus Training School) by Mrs. Brite. The cards were presented in the indicated order. The stories were tape recorded and later transcribed with two copies of each protocol being made.

Drawing from clinical experience, common sense, and research information, Mrs. Brite and the investigator together subjectively scored part of the 26 protocols and began to establish both a method of approach to the stories and a list of "helps" (guide lines) that would help differentiate between the personality variables. The "helps" consisted of ideas, words, phrases, etc., which the scorers felt reflected a certain attitude. These were compiled into three lists (aggression, depression, and feelings of inferiority) which served as reference guide lines in the evaluation of subsequent stories. Throughout the three developmental phases--and, in fact, throughout the final

test scoring phase--the list of "helps" was expanded and refined. In some instances "helps" items were given set rating scores. As an example, in the rating of aggression, if a pupil's story stated that a person had been murdered or killed, the variable of aggression was given the rating of 5 (100 per cent aggression). If a pupil stated that the person "felt like" killing or murdering someone, the rating was 3 (50 per cent aggression). Specific rating values were assigned to only a few of the more common "helps" statements and served to provide a structured frame of reference that proved to be of immense help to both raters. Assigning specific scores to particular "helps" items did not limit the subjective interpretation of the story, as the rater was free to decide how he felt the particular concept was being utilized. It became apparent during the course of scoring large numbers of stories that the raters themselves became concerned as to their ability to maintain a consistent frame of reference in relation to a particular personality variable. The assigning of a specific value to certain "helps" items offered a concrete check point which helped to assure the rater that he was still maintaining the proper orientation.

In many instances, a particular "helps" statement suggested several interpretational alternatives. For example, the statement, "The boy decides to run away," could indicate an aggressive act. Conversely, if the context connotation suggested that this was a withdrawal reaction, the statement might then be interpreted as an indication of inadequacy or inferiority. The context clues always decided the final rating score and gave the rater complete subjective latitude in deciding which score was to be given. The lists of "helps" follow.

HELPS

VARIABLE--AGGRESSION

<u>"Helps" Item</u> <u>(statement or inference)</u>	<u>Score</u>	<u>Words and Expressions</u> <u>Suggestive of Aggression</u>
Murdered or killed	5	Broke something Shoot
Hostile impulse		Knock
a. demonstrated	4	Hurt
b. felt like	3	Hit
Mad	3	Grab Throw Push
Refused to see obvious detail in picture	3	Naughty Do purposeful wrong Violate rules
Refusal to tell story	3	Defy authority Blame others
Card rejection	3	Others blame him Run away
Confused story	1	Run I don't know
Unscorable story	1	Win every day Need to win daily
Hate	2 or 3	Better straighten up Got a good talking to Punishment for wrong- doing
Lie		Escape
Runs away to punish parents		Forget, purposefully Don't know how Stealing
Always late, opposition		Scolding
Can't think, not even going to try, and other indications of resistance		Not getting lessons done Gripes and complains, plus projection or displacement
Something that happened a while ago, indifference		Fight Swear Call names
Sleep		Hate
Melodramatic stories, if inappro- priate		
Retribution		
Wishful thinking		
Unpleasant job assigned (qualify)		

Outgoing or positive aggression

Rejection of aggression or story
reconstructed to deny hostility

Reconstitution decreases the score;
not simply a statement, but a real
change in attitude; not denial,
but more of a reform

Denial of aggression (increases
score)

Won't see what is obvious

VARIABLE--DEPRESSION

<u>"Helps" Item (statement or inference)</u>	<u>Score</u>	<u>Words and Expressions Suggestive of Depression</u>
Suicide	5	Probably I guess
Poverty of response	3	Maybe Could be
Disappointment	3	Indifferent attitude Hopelessness
Unscorable story	1	Sad Tired
Three "D's" of depression		Or something
1. Deprivation		Being denied
2. Denial		Discouraged
3. Disappointments		Worried
Cry in disappointment		Hurt Reactive depression
Hopelessness		Crying Inability to deal with
Being left out		
Sleep		
Chronic inability to compete		
Duration; has to do it a long time		
Parent's not taking care of, neglect		
Lonely		

Something is always this way

Die, deprivation

Feels sorry for self

VARIABLE---INFERIORITY FEELINGS

<u>"Helps" Item</u> <u>(statement or inference)</u>	<u>Score</u>	<u>Words and Expressions Sug-</u> <u>gestive of Feelings of</u> <u>Inferiurity</u>
Being laughed at	4 or 3	Trying
Ridiculed		Cry
Lonely		Ask questions can't answer
Embarrassed (context)		Worry
Not able to compete		Ashamed
Powerless to act		A dummy (person)
Scared		Staring
Wishful thinking		Kids laughing at him
Never on time		Teasing
Escape, run away		Compliancy (dependent upon context)
Won every day		Make fun of
I don't know		Laugh at
Need for punishment		Fail test
Day dreaming and wishing		Flunks test
Sick		
Failed		
Unaccepted		
Feels sorry		

A 5-point rating scale was established, as follows:

<u>Score</u>		
1	0 per cent	Normal
2	25 per cent	Some more than normal
3	50 per cent	Fair
4	75 per cent	Considerable
5	100 per cent	Extreme

In general, manifest content was scored; but in cases where doubt existed, the scorer was free to use his subjective impression based on inferences drawn from the total TAT protocol.

The procedure for evaluating each story was as follows:

1. Each story was presented to each rater in the same order.
2. Each story was read twice by each rater.
 - a. The first reading involved getting the total gist of the story and possibly assigning numerical scores for each variable.
 - b. The second reading required synthesizing the diagnostic impression and rating each variable according to the 5-point scale.
 - (1) As an example, a rater might read a story and assign the ratings: aggression, 1; depression, 4; inferiority, 2, which might suggest that the story indicated normal aggression, considerable depression, and more than a normal amount of inferiority.
3. Stories were rated on the 5-point scale.

4. Scores were recorded on a prepared score sheet which allowed for the tabulation of the scores for each story and a final sum score for each personality variable.

After the initial list of helps and scoring methodology had been established, each rater scored protocols independently, and frequent comparisons and discussions were held to refine methodology, expand the list of "helps," and share personal insights. As a check on the reliability of the scoring method, two inter-rater reliability coefficients were computed and are as follows:

Table 1. Inter-rater reliability coefficients, Edith Bowen School sample (Spearman-Brown Correction applied)

Number	Variable		
	Aggression	Depression	Inferiority
7	.81	.96	.91
12	Not now available	.68	.30

The results indicated a need for further refinement and the second phase of development was undertaken.

Second phase

In the second phase of development, Michigan Picture Test card 2 was discarded, as it didn't seem to discriminate between the variables being studied. The test was then administered by the standard method to 144 fifth-grade pupils in District R and District A (40 boys and 33 girls in District R; 40 boys and 31 girls in District A). The

stories were recorded on tape and later typed in duplicate copy. All identifying information except the sex of the pupil was removed from the protocols and the raters had no knowledge of the pupil's district or grade level. Protecting the anonymity of the pupil was considered basic procedure in reducing scoring bias. The protocols were randomized by shuffling and then scored in units of 15 to 30 per week, according to the established method and utilizing the "helps" that had been accumulated. During the scoring of these tests, as was the practice during all three phases of development, the "helps" lists were continually revised as new meaningful material was elicited. During the process of scoring, the raters met at least weekly (oftener if the opportunity arose) to discuss methodology, revise the "helps" lists, and talk over any particular problems that may have been encountered.

In an effort to clarify scoring problems, the raters frequently would select a previously scored test which had presented particular scoring problems and discuss the scoring procedure for that protocol in detail. In spite of disagreements, original scores were never altered and only previously-scored protocols were discussed. Test protocols were always rated independently by each rater and comparisons always followed the summation and recording of the original scores.

The weekly meetings are considered essential to the scoring method, as the basis of orientation is easily lost over fairly long periods of time. Also, it was noted that a particular rater may, due to particular problems in his own life adjustment, become biased in his ratings. This weekly meeting served to point this up and functioned as a control over individual attitude sets. During the course of scoring,

frequent inter-rater reliability coefficients were computed as a further safeguard against individual bias and, as the method was still in a stage of development, to check the reliability of the results. The cumulative reliabilities are reported in Table 2.

Table 2. Second phase accumulative inter-rater reliability coefficients (Spearman-Brown Correction applied)

Number	Variable		
	Aggression	Depression	Inferiority
12	.81	.66	.92
18	.93	.17	.78
12 + 18 = 30	.88	.28	.86
18	.94	.60	.45
18 + 30 = 48	.91	.39	.60
29	.90	.93	.95
29 + 48 = 77	.89	.86	.88
33	.96	.91	.86
33 + 77 = 110	.92	.89	.90
11	.91	.83	.86
11 + 110 = 121	.94	.87	.93
23	.93	.68	.85
23 + 121 = 144	.94	.86	.90

It will be noted that after scoring 144 8-story protocols, the inter-rater reliability coefficients across the three personality variables clustered around .90. This was considered satisfactory, and the scoring method was considered reliable enough to apply to the present study, or phase three.

Third phase

Phase three involved several procedural changes which are outlined in detail in the chapter on procedures, but for purposes of clarification, they will be presented here in brief form.

The major procedural changes were: (a) Several of the stimulus pictures were changed, (b) the pictures were presented to groups of sixth-grade pupils by means of an opaque projector in a semi-darkened room, and (c) pupils were required to write 7-minute stories to each of the seven TAT-type pictures.

It was necessary to change some of the initial stimulus pictures because of the change in method of test administration and because of a desire to have as many pictures designed to stimulate responses from girls as from boys. The selected TAT-type pictures used in the final phase of the experiment were, in order of presentation, as follows:

	<u>Test Card No.</u>	<u>Description</u>
1.	TAT 1	Boy sitting at desk with violin.
2.	MPT 3	Classroom situation.
3.	TAT 3 BM	A boy huddled by a coach with a revolver at his side.
4.	MPT 8 G	A young girl sitting at a school desk staring into space.
5.	TAT 7 GF	A young girl with an older woman.
6.	MPT 8 B	A young boy with his head resting on his hands staring into space.
7.	MPT 12	Blank

Card 1 was selected for its generality. Pupils of both sexes appeared to be able to identify with the figure. Card 2 was included to stimulate responses pertaining to the school situation. Card 3 is quite vague and could elicit responses pertaining to any or all of the personality variables being studied. Cards 8 G and 8 B depict similar situations, but are sex oriented and were included in an effort to present pupils with equivalent stimuli. These cards are vague in their implications and could allow a wide range of fantasy projection. Card 7 GF was included to depict a common situation with which most children can identify, but also because card 3 BM depicted a male figure and a picture depicting a girl in a similar relationship was desired. The blank card was included both as an experiment and as a means of offering the pupil an unlimited range of fantasy expression. In practice, the blank card proved to be a distracting influence and a source of frustration to many of the pupils--especially the low-ability pupils. It is felt that it should not be included in further research with children of this age unless some special effect is desired.

Having pupils write stories to TAT pictures instead of recording their responses according to the standard procedure considerably shortened the length of the stories elicited, and offered the raters fewer cues from which to derive a score. This obstacle was dealt with on the basis of projective technique rationale. The writer felt that if the TAT is an instrument capable of eliciting meaningful samples of a pupil's personality, and if a pupil really projects into TAT stories the sum total of his feelings, then a story of any reasonable length would contain the elements of his personality adjustment.

The method was applied to a sample of 338 TAT-type story protocols obtained by the method described in phase three. The resulting inter-rater correlation coefficients were obtained: aggression, .91; depression, .70; and inferiority, .73. Table 3 presents these results.

Table 3. Phase three inter-rater reliability coefficients (Spearman-Brown Correction applied)

Number	Variable		
	Aggression	Depression	Inferiority
338	.91	.70	.73

The reduction in inter-rater reliability r 's for the variables of depression and inferiority is probably due to statistical regression. In the final use of the scoring method, several variables (stimulus cards, method of administration, and length of story) were changed; and the statistical tendency with the addition of variables is to regress scores towards the mean. The variable of aggression did not demonstrate regression, probably, because aggression is a much easier concept to evaluate than depression or inferiority in any sample. The resulting inter-rater correlation coefficients are considered adequate, as they approximate those reported in the literature previously presented.

Suggestions Concerning the Use
of This Scoring Method

In keeping with Eron, Terry, and Callahan's (1950) inferences, it is desirable that a scoring method contain enough structure to allow

subsequent raters to utilize the method with sufficient reliability to be useful. During the process of developing this scoring method, several desirable methodological "helps" were noted and should be practiced in any further attempts to apply this scoring method:

1. Raters should rate enough sample stories to obtain the desired inter-rater reliability before rating the experimental sample tests.

2. Test protocols must be anonymous.

3. The list of "helps" should be closely adhered to, and new material should be added to the "helps" lists as it become available. (This should be done in a conference with the other rater(s).)

4. Protocols should be scored as rapidly as possible. This is important, as inter-rater reliability appears to be a function of time.

5. Conferences between raters should be frequent (preferably twice weekly).

6. Cumulative inter-rater reliability coefficients should be calculated frequently; and when discrepancies appear, conferences should be more frequent.

7. Statistical equivalence should be practiced. The sample should be randomly selected from the population and the selection of tests for scoring should be random. As has been stated, environmental influences affect the ratings of individual raters, and every effort should be made to reduce statistical bias by maintaining rating reliability and consistency. Several things can be done to reduce the effects of environmentally-induced rater bias:

- a. The number of protocols scored at each sitting should be small.
- b. Each protocol should be rated by each rater at approximately the same time.
- c. Raters should work in the same environmental surroundings.
- d. Raters should not be allowed to score for more than one hour without a rest; fatigue was found to have a profound detrimental effect upon inter-rater reliability.

Used in keeping with these guides, and applied to the variables of aggression, depression, and feelings of inferiority, this scoring method should be useable by an individual who has a reasonable knowledge of TAT principles and interpretation. It is felt that the method could also be applied to other personality variables if the necessary "helps" lists were developed.

METHODS AND PROCEDURES

This study employed a balanced design in which randomly selected, equal samples of sixth-grade pupils were taken from three school pupil populations--superior, average, and below average--under two classroom grouping conditions: homogeneous (ability grouped and heterogeneous (random grouped)).

The sample was selected from two adjacent school districts which are situated contiguous to each other. Schools within each district were selected which were most comparable in terms of socio-economic status and rural-urban environment. The pupils were in most respects comparable across districts. District R (control group) employed a heterogeneous grouping system, and District A (experimental group) employed a three-level homogeneous ability grouping system. District A had been using the homogeneous grouping system for two years, data for this study being collected during the third year, and the pupils selected for the study had been in the grouping situation for about that length of time.

Method of Sample Selection

As District A was already practicing ability grouping, it was necessary to establish a suitable criteria for selecting an equivalent sample from the non-grouped District R.

In District A, pupils were assigned to the appropriate ability group according to the following procedure:

1. Prior to the administration of the achievement test, the teacher would be requested to complete an official form in which he would recommend that each child in his class be placed in the average, developmental, or accelerated section for the coming year. This form should be completed prior to testing so that a teacher's judgment will not be unduly influenced by the child's performance on the California Achievement Test (CAT).

2. After the achievement tests have been scored, a roster for each sex and grade will be prepared listing the name of each child in order of his total score on the CAT. For example, the highest fourth grade boy would be placed first on the boys' list and so on until all children have been listed. Boys and girls would be placed on separate lists.

3. In order to set up the developmental section, the lowest ten boys and lowest ten girls on a basis of CAT total grade placement will be chosen. This will give an initial group of 20 children. If any children in this group are more than .5 grade placement points (one-half year) above the rest of the group, these children will be placed in an average group. Every effort should be made to keep the ratio of boys to girls approximately equal. This step may be altered to some degree by the principal to fit the situation, but in no case will the initial group exceed 20 students.

4. The remaining children will be divided into the number of accelerated and average sections to be made up. For example, if the school had 120 fourth grade children approximately the lowest ten boys and lowest ten girls would be placed in the developmental section and the remaining 100 would be divided into three sections of about 30 to 35 each.

5. To make up the accelerated section the top 15 boys and the top 15 girls would be tentatively chosen.

6. This would leave sufficient children to make up two average groups. These average groups would be chosen by placing the odd numbered boys and girls in one section and the even numbered in another. This procedure would result in the average children being heterogeneously grouped within the average range.

7. In schools having five or more sections in a given grade, two developmental sections or two accelerated sections may be formed if the CAT distribution, as compared with other schools in . . . County, justifies this arrangement.

8. After this tentative grouping is completed, children in each group who are within .5 grade placement units of the next group would be checked and adjustments based on teacher recommendations would be made. For example, if a child is at the high boundary of the average group and the teacher recommendation is to place this child in the accelerated group, the child would be placed in the accelerated group. These adjustments would not be made except in case of borderline children.

9. In cases where the teacher's judgment disagrees radically with the test score (teacher recommends developmental placement, test recommends accelerated placement or vice versa), the scoring will be checked and if correct, the child will be retested with another form of the CAT.

a. If the retest agrees with the first test, the child will be placed in accordance with the test results.

b. If retest agrees with teacher, the child will be placed in accordance with the teacher's opinion.

c. If retest does not agree with either the initial test or teacher opinion, the child will be placed in an average group.

10. Records would be kept of children on whom the test and teacher disagreed radically and if a sufficient number of the cases occurred, a special study of these pupils would be made.

11. Children transferring to . . . County schools from other districts should be tested during their first day in school. Their total CAT grade placement should be corrected for the time of the year and they should be placed in an accelerated, average, or developmental section based on the corrected CAT grade placement total. If borderline on the CAT, the child will be placed in an average section.

Using this system of classification, District A pupils were given the California Achievement Tests (CAT) early in May, 1960, and were assigned to their respective ability groups for the school year beginning 1961. In April, 1960, Utah State University, as part of a larger study, administered the Sequential Test of Educational Progress (STEP) to all fifth-grade pupils in both District A and District R. As District A had already identified and grouped the pupils for the coming year, 1961, the first step in establishing a criteria by which to select an equivalent District R sample was to obtain the mean STEP score for each level (superior, average, and developmental) ability group in District A. Cut-off scores between each ability level were then calculated by the split-mean method.

$$\frac{\text{Mean Accelerated} + \text{Mean Average}}{2} = \text{Cut-off score between average and accelerated}$$

$$\frac{\text{Mean Developmental} + \text{Mean Average}}{2} = \text{Cut-off score between average and developmental}$$

The cut-off scores determined are presented in Table 4.

As the District A pupils were already grouped within their respective schools, it was a simple matter to select enough schools and test enough classes to obtain the desired number of cases in each group. In general, as an attempt to maintain random sampling, an effort was made to test all of the sixth-grade pupils in each of the selected schools. However, as two of the larger schools had several average groups, only one average section was tested in each of these schools. All superior and developmental groups were tested in each of the selected District A schools.

Obtaining the desired sample in the District R schools presented other problems. As the pupils were grouped heterogeneously, it was sometimes necessary to test an entire class in order to obtain data on one or two developmental or superior children. As a result, it was necessary to test 707 District R pupils in order to obtain a sample of 25 random-grouped developmental girls.

Development of the Testing Procedure

Three tests (a modified version of the Thematic Apperception Test (TAT) the Test Anxiety Scale, and the Children's Manifest Anxiety Scale) were used in the investigation of the four personality variables aggression, depression, feelings of inferiority, and anxiety. Seven TAT-type pictures were shown to classroom groups of sixth-grade pupils in two adjacent school districts. The pupils were required to write 7-minute stories to the stimulus pictures.

The group method of TAT administration is a newer type of approach and little research had been done as to the effect of this type of TAT

administration on young children. Johns (1960), in a study of achievement motivation, projected TAT pictures to groups of sixth-grade pupils and felt that pupils of this age could be expected to write five-minute stories to as many as ten TAT pictures without extreme fatigue or loss of interest.

In preparation for the testing for the present study, it was decided to conduct a trial test on a group of pupils similar to those to be used in the final study. The purpose of the trial test was to (a) try out the three tests and determine the order of presentation, (b) test the utility of projecting TAT pictures by means of an opaque projector in the usual classroom situation, (c) determine how many TAT pictures should be presented, (d) determine how much time pupils needed in order to write their stories, (e) obtain a sample of written stories that might be scored by the developed method (no one really knew how adequate the written stories of sixth-graders would be), and (f) have an opportunity to try out and refine administrative procedures.

One of the schools in District A was selected for the trial run, and the superior, average, and developmental sections were given the experimental test battery.

It had been decided that, whenever possible, pupils should be tested in their own classroom situation and that students would never be grouped for the purpose of administering the tests. This was done in an effort to reduce the possibility that intervening variables such as anxiety, hostility, or the effects of a new situation would influence the results. In some cases, it was impossible to get a particular classroom dark enough to use the opaque projector. Those classes had

to be tested in the audio-visual room, auditorium, or lunchroom. This situation occurred in both districts, and the effects were probably compensatory. Also, pupils in both districts were used to being tested under these conditions, so the over-all effect was probably minor. In future studies using this procedure, it might be better to plan on using the usual testing situation utilized by the school in order to avoid inconsistencies.

The results of the trial test indicated that in general sixth-grade pupils could write adequate stories to as many as seven TAT pictures, but that past seven or eight the pupils became restless and disinterested. Five minutes writing time was found to be too short for the superior pupils, and many of the average pupils expressed a need for a "little more time." Most of the developmental pupils were through in five minutes, but a few of those who really wanted to try needed more time. Ten minutes writing time either fatigued most pupils at each level or left time for "monkey business" between card administrations. Seven minutes writing time was selected and proved satisfactory in all situations, although occasionally a pupil would not have an opportunity to complete his story.

As the TASC instructions contain an introduction to the test situation, it was given first. This was followed by the TAT (seven cards), and the CMAS was given last. As the test battery was long, and as the CMAS is a yes-no-type questionnaire which is completed by the pupil individually, it was possible for the examiner to collect and assemble most of the test materials while the children were completing the CMAS, thereby saving time. It was found that the test battery could be given

in an hour and a half (allowing for a three-minute break between TAT cards 4 and 5) if one proceeds rapidly. By adhering strictly to the time limits and the established administrative procedure, it was sometimes possible to test two different classes in a single morning.

All of the trial test sessions were tape recorded, and an effort was made to structure the final test instructions so that most of the questions commonly asked by the pupils were answered in the instructions. After each trial test situation, the class was encouraged to offer suggestions that would make the test better and to describe their feelings about this kind of test. In general, most of the children considered the test a new experience and many stated that they "wished more tests were like this."

The quality and quantity of the obtained TAT stories proved to be adequate for scoring, and the following procedure was used in the final collection of the experimental data.

As has been mentioned, the length of the test battery (one and one-half hours) necessitated a strict adherence to the administrative procedure. In each situation the transition from the classroom to the test situation was made within a few minutes. It was the practice to meet the classroom teachers before the scheduled test period and instruct them as to their role in the procedure. As teachers were always required to leave the room during the test period, their role was to introduce the examiner: "This is Mr. Swaner, from Utah State University, who is doing some very important research," state that they must leave the room during the test period, and leave. Testing began immediately after the teacher left the room.

(See Appendix for detailed description of the test administration procedure and instructions.)

Methods of Scoring

After all of the data had been collected, the protocols were separated into their twelve respective groups (described in Method of Statistical Analysis which follows) and each protocol was checked for completeness. Occasionally a pupil would fail to write a story to a particular picture stimulus, write an incomprehensible story, or in some other way distort the stimulus response. These protocols were rejected as unscorable. Also, the CMAS (L) scale was utilized as a method of screening protocols, and all pupils who obtained an (L) score of over 7 were eliminated from the sample. During the course of testing, the pupils were observed carefully, and the test protocol of any pupil who would obviously not make an effort was of course eliminated from the sample. This situation occurred infrequently, but was present. Also, on the developmental level, several pupils could neither read nor write, and these protocols were rejected; however, the pupils were allowed to remain in the test situation in an effort to maintain class homogeneity. In most cases these pupils were identified prior to the test situation, and no effort was made to encourage them to respond, although they dutifully attempted each test.

The application of a rigid screening procedure resulted in the rejection of a considerable number of test protocols, and it is recommended that any future studies utilizing this method of data collection and screening allow for at least a 30 per cent sample reduction in

scorable test protocols. Following screening, 30 protocols were randomly selected from each of the twelve groups (with the exception of the sample of developmental girls in both districts, in which cases all of the scorable protocols were utilized because of the limited number available) and the selected protocols were randomized into one group.

An individual not associated with the scoring procedure removed all identifying data except age and sex from the test protocols, and the protocols were presented to the raters in units of twenty, for scoring. The protocols were then scored by the method previously described.

Method of Statistical Analysis

The study utilized a simple analysis of variance design in investigating the significance of the differences among the means of the 12 different conditions. The 12 conditions represented the district, level, and sex of pupils in each grouping system. Conditions 1 through 6 designate the heterogeneously grouped pupils in District R; conditions 7 through 12, the homogeneously-grouped pupils of District A. Their respective order is as follows:

<u>District R</u>	<u>District A</u>
1. Dis. R, superior, boys	7. Dis. A, superior, boys
2. Dis. R, average, boys	8. Dis. A, average, boys
3. Dis. R, developmental, boys	9. Dis. A, developmental boys
4. Dis. R, superior, girls	10. Dis. A, superior, girls
5. Dis. R, average, girls	11. Dis. A, average, girls
6. Dis. R, developmental, girls	12. Dis. A, developmental, girls

The 12 conditions presented 24 possible district, level, and sex comparisons for each personality variable (permitting the manipulation of one variable), which were then tested for significance according to the standard analysis of variance technique. The resulting significant variance ratios (F) were tested for separate differences by the standard t test (Garrett, 1959, pp. 280-87).

The 24 comparisons made were as follows:

District comparisons:

- | | |
|--------------------------------|--------------------------------------|
| 1. Dis. R, superior, boys | vs. 7. Dis. A, superior, boys |
| 2. Dis. R, average boys | vs. 8. Dis. A, average, boys |
| 3. Dis. R, developmental boys | vs. 9. Dis. A, developmental boys |
| 4. Dis. R, superior, girls | vs. 10. Dis. A, superior, girls |
| 5. Dis. R, average, girls | vs. 11. Dis. A, average, girls |
| 6. Dis. R, developmental girls | vs. 12. Dis. A, developmental, girls |

Level comparisons:

- | | |
|-----------------------------|--------------------------------------|
| 1. Dis. R, superior, boys | vs. 2. Dis. R, average, boys |
| 1. Dis. R, superior, boys | vs. 3. Dis R, developmental, boys |
| 2. Dis. R, average, boys | vs. 3. Dis. R, developmental, boys |
| 4. Dis. R, superior, girls | vs. 5. Dis. R, average girls |
| 4. Dis. R, superior, girls | vs. 6. Dis. R, developmental, girls |
| 5. Dis. R, average, girls | vs. 6. Dis. R, developmental, girls |
| 7. Dis. A, superior, boys | vs. 8. Dis. A, average, boys |
| 7. Dis. A, superior, boys | vs. 9. Dis. A, developmental, boys |
| 8. Dis. A, average, boys | vs. 9. Dis. A, developmental, boys |
| 10. Dis. A, superior, girls | vs. 11. Dis. A, average, girls |
| 10. Dis. A, superior, girls | vs. 12. Dis. A, developmental, girls |
| 11. Dis. A, average, girls | vs. 12. Dis. A, developmental, girls |

Sex comparisons:

- | | | |
|--------------------------------|-----|----------------------------------|
| 1. Dis. R, superior, boys | vs. | 4. Dis. R, superior, girls |
| 2. Dis. R, average, boys | vs. | 5. Dis. R, average, girls |
| 3. Dis. R, developmental, boys | vs. | 6. Dis. R, developmental girls |
| 7. Dis. A, superior boys | vs. | 10. Dis. A, superior, girls |
| 8. Dis. A, average, boys | vs. | 11. Dis. A, average, girls |
| 9. Dis. A, developmental, boys | vs. | 12. Dis. A, developmental, girls |

Following the collection and scoring of the data, the tests were grouped into their appropriate conditions and test scores were recorded on hand data cards and then programed onto IBM cards. The IBM then computed the N, sums of scores ($\sum X$), and the sums of scores squared ($\sum X^2$) for each of the test variables across the 12 conditions. The scores obtained from the TASC, CMAS, and CMAS (L) scales were treated in the same manner. The analysis of variance, F and t tests, was computed by hand, using the summarized IBM information.

Inter-rater reliability coefficients were computed from the following formula, Coefficient of Correlation, calculated from obtained scores (Garrett, 1959, p. 143).

$$r = \frac{N\sum XY - \sum X \times \sum Y}{\sqrt{[N\sum X^2 - (\sum X)^2](N\sum Y^2 - (\sum Y)^2)}}$$

The procedure utilized in calculating the 24 group comparisons was uniform except in two cases, which were as follows:

1. Across the variable of inferiority, no significant sex differences were found. This offered the opportunity of combining boys and girls at each level in their respective districts, thereby

RESULTS AND DISCUSSION

The primary purpose of this study was to investigate the differential effects of heterogeneous versus homogeneous grouping practices on the personality structure of the pupils in each situation across the personality variables of aggression, depression, inferiority, and anxiety. As a secondary purpose, the research design endeavored to develop and test a partially objectified, inter-rater type rating scale method of scoring Thematic-Apperception-Test-type stories according to the personality variables of aggression, depression, and feelings of inferiority.

The hypotheses stated in the introduction of this paper will be reviewed here, the findings pertinent to each will be given, and the findings will be discussed.

Scoring of TAT-Type Stories

Hypothesis No. 1 stated: "Thematic-Apperception-Test-type stories can be scored by a five-point rating scale with sufficient accuracy to permit inter- and intra-group comparisons on the personality variables of aggression, depression, and feelings of inferiority."

Findings

Two raters independently scored 338 seven-story TAT-type protocols obtained from sixth-grade pupils, according to the previously-described scoring method. The resulting inter-rater correlations,

corrected by the Spearman-Brown Prophecy Formula (Garrett, 1959, p. 343) are presented in Table 8.

Table 8. Final corrected inter-rater reliability coefficients

Variable	Size of sample	Coefficient of reliability	Spearman-Brown Prophecy Formula corrected
Aggression	338	.84	.91
Depression	338	.54	.70
Inferiority	338	.57	.73

The corrected inter-rater r 's were comparable with those obtained by other rating methods (Harrison and Rotter, 1945; Hokanson, 1958; Fine, 1955b) and were considered adequate. Inspection of the corrected reliabilities suggested that aggression was rated more effectively than either depression or inferiority. The latter two variables required a great deal of subjective judgment, which may have introduced error.

District Comparisons

Hypothesis No. 2 stated: "In relation to the selected personality variables (anxiety, aggression, depression, and feelings of inferiority) there will be no significant differences between similar pupils in the two districts, except for the following:

- a. Superior and developmental pupils in the heterogeneously-grouped classroom will obtain higher mean scores across the

variables of aggression and depression than similar pupils in homogeneously-grouped classrooms as an effect of frustrations imposed by the restrictions of a school situation that is geared to the needs of the average pupil.

b. Homogeneously-grouped superior pupils, as a result of intra-group competition and fears associated with losing social status if not successful, will obtain higher Manifest and Test Anxiety mean scores than heterogeneously-grouped superior pupils.

c. Homogeneously-grouped developmental pupils will, as a result of being classified as 'dumb,' obtain higher mean scores on the personality variable 'feelings of inferiority,' than similar heterogeneously-grouped pupils."

The findings and interpretation of the over-all district comparisons will first be made, and then the sub-groups of the hypothesis will be presented.

Findings

A total of 36 comparisons were made between similar pupils in District A (homogeneously grouped) and District R (random grouped). Of these, only six differences were significant beyond the 5 per cent level. These differences (as summarized in Table 9) were as follows:

1. District A superior boys demonstrated more aggression than District R superior boys.
2. District R average girls obtained higher depression scores than average girls in District A.
3. District R developmental girls demonstrated more depression than developmental girls in District A.

Table 9. Significant district differences

District comparisons	Personality variables					
	Aggression	Depression	Combined Inferiority	TASC	CMAS	Total CMAS (L)
<u>Superior ability level:</u>						
Boys:						
District A vs. District R ^a	.01 ^b					.01 ^c
Girls:						
District A vs. District R						.01
<u>Average ability level:</u>						
Girls:						
District R vs. District A		.05				.01
<u>Developmental ability level:</u>						
Girls:						
District R vs. District A		.05				

^aThe first group listed obtained the higher score; higher scores are considered less desirable.

^bLevel of significance.

^cReversed relationship; latter group listed obtained higher score.

4. District R superior boys obtained higher CMAS (L) scores than District R superior girls.

5. District A superior girls obtained higher CMAS (L) scores than District R superior girls.

6. District R average girls obtained higher CMAS (L) scores than District A average girls.

Interpretation

The rationale underlying this primary hypothesis is that children bring to the school situation a fairly well-developed, culturally-determined personality--learned primarily from their parents. This personality is fairly well fixed by the sixth grade in school, and the school situation is merely a sounding board against which they try out their learned patterns of behavior. In other words, a child who is aggressive, anxious, depressed, or experiencing feelings of inferiority in the school situation has learned this characteristic method of response primarily in the home situation. As the pupils in this study came from a fairly homogeneous cultural setting, and as personality traits seem to be normally distributed across a given population, it seems reasonable that one would find an equal amount of any particular variable in any representative group of individuals.

If the school is considered to be merely a situation in which a pupil responds with his learned behavior, conditions within the schools would have to be drastically different to elicit extreme differences in the adjustments of any particular individuals. In other words, by the time he reaches the sixth grade an individual has

established characteristic attitudes and personality patterns which will not vary considerably as a function of any particular situation; i.e., an anxious child is an anxious child anywhere.

In two random samples of a normal population, therefore, one would expect to find as many children manifesting a particular personality characteristic in one sample as in the other. These assumptions are in accord with the work of Sarason et al. (1960), in that parental influence appears to be the primary influence upon the child's school adjustment.

The present findings supported these assumptions in that none of the personality variables investigated in this study demonstrated any consistent, significant between-district differences.

Aggression and depression

In regard to the first sub-group of this hypothesis, a review of the pertinent literature suggested that the superior and developmental pupils in the heterogeneously-grouped classroom might obtain higher mean scores across the variables of aggression and depression than similar pupils in homogeneously-grouped classrooms, as an effect of frustrations imposed by the restrictions of a school situation that is geared to the needs of the average pupil.

Findings. The only finding in keeping with the hypothesis was that District R developmental girls demonstrated significantly (5 per cent level) more depression than developmental girls in District A. An additional finding was that District R average girls obtained significantly (5 per cent level) higher depression scores than average

girls in District A. In relation to the variable of aggression, District A superior boys demonstrated more aggression than District R superior boys (significant at the 1 per cent level of confidence).

Interpretation. This hypothesis was based on the assumption, derived from Mathias' (1959) study, that the wide spread of individual differences within the heterogeneously-grouped classroom would frustrate both the bright and the slow pupils and cause a greater amount of aggression and depression in both groups. In relation to aggression, the present study indicates an opposite result and applies only to boys--with the bright boys grouped on ability level demonstrating a significantly greater amount of aggression than similar pupils in the supposedly frustrating heterogeneous classroom situation. The fact that the relationship is inverse, and does not hold for the developmental level, casts serious doubt upon Mathias' interpretation of his findings.

In relation to depression, the findings are in keeping with Mathias' findings in that the developmental (low I.Q. and achievement group) students demonstrated more depression (5 per cent level) in the heterogeneous group situation than in the homogeneous group situation. However, the relationship applies only to girls and extends upward to include only the average group, which suggests that both average and developmental girls demonstrated more depression than similar pupils grouped for ability. Mathias' inference was that the average group would demonstrate less depression than either of the extreme groups. It is possible that girls of average or low ability, grouped on the basis of ability, feel more comfortable, accepted, and have more opportunities for success than similar pupils in the heterogeneous

group situations wherein direct competition with superior pupils may limit their opportunities to demonstrate adequacy. However, the work of Luchins and Luchins (1948) suggested that pupils in the lower ability groups should feel more depressed because of feelings of discrimination in an ability-grouped situation.

Mathias' and Luchins and Luchins' findings applied to both girls and boys as a group. However, the present findings suggest that a sex differential is present with superior boys in the homogeneously-grouped situation demonstrating more aggression than superior boys in the heterogeneously-grouped situation, and average and developmental girls in the heterogeneously-grouped situation demonstrating more depression than similar pupils in the homogeneously-grouped situation. This suggests that any inferences made as to the effects of a grouping practice with any particular group of school children may have to consider the sex of the pupils, as well as level of academic achievement.

This finding could have relevant applications to the current trend toward grouping pupils as a means of facilitating the development of special abilities through school enrichment programs. If one considers the ultimate effects of aggression and depression detrimental to the development of the pupil, the present findings suggest that grouping superior boys may be detrimental to their personality development, and that superior boys might profit most from a heterogeneous group situation. Likewise, the findings might suggest that average and developmental girls would develop more adequately in a homogeneous grouping situation. It is evident that the findings raise more questions than they answer, but in keeping with the present emphasis in education

upon the recognition of individual and group differences, the findings suggest what might be a fruitful area for further research.

The demonstration of a sex differential in grouping practice places strong doubts upon the findings of Mathias and Luchins and Luchins, neither of whom utilized a control group in their study. The matter of whom, what, when, and how to group appears to be a very complex area of investigation.

Anxiety

In regard to the second subgroup of this hypothesis, (b), the assumption was that homogeneously-grouped superior pupils would react to the narrower range of intra-group competition with an increase in anxiety, centered around the fear of losing social status if not successful.

Findings. There were no significant district differences in anxiety scores on the TASC or CMAS tests.

Interpretation. The results of the study reported by Luchins and Luchins (1948) suggested that homogeneous grouping practices seemed to help create a sort of caste system in the school, which had the profound effect of making superior pupils overly concerned with the stability of their status. The strength of their concern was evident when some superior pupils expressed the feeling that they would rather take the chance of failing in the bright group than be in any way associated with the average or dull groups. The fear associated with losing status might be expected to produce considerable anxiety. Engle (1938)

also suggested that bright students may develop increased anxiety as a result of increased expectations and a feeling of being different from their peers. The present findings suggest that the grouping factor is not the significant variable. However, level and sex differences were indicated and suggest that inter-group and intra-group relationships within the particular grouping system are more important than the effects of grouping or not grouping. The differences will be treated more completely in regard to level and sex comparisons.

Feelings of inferiority

In regard to the third subgroup (c) of the second hypothesis, it was hypothesized that homogeneously-grouped developmental pupils would react to being classified as "dumb" and demonstrate higher mean scores on the personality variable of feelings of inferiority than similar heterogeneously-grouped students.

Findings. There were no significant sex or district differences. As there were no significant sex differences, boys and girls at each level were grouped for the purpose of increasing the size of the sample N in each group, and the combined groupings were tested for significance. The combined grouping did not demonstrate any significant changes.

Interpretation. The hypothesis that homogeneously-grouped dull students would react to the stigma of "being dumb," as purported by Luchins and Luchins (1948), with increased feelings of inferiority, was completely refuted. The fact that there were no significant district differences suggested that the method of grouping practice

was probably not the significant factor in the production of feelings of inferiority in this sample of sixth-grade children.

The absence of a relationship between the hypothesized effect of grouping practice and feelings of inferiority and the present findings could lead to a good deal of speculation as to the reasons. Common sense might suggest that children labeled as "dumb" might experience some reduction in level of self-confidence, self-concept, or feelings of adequacy. The fact that pupils in heterogeneous grouping situations demonstrated no sex or level differences in feelings of inferiority suggests that being labeled "dumb" may have effects other than reducing feelings of adequacy. The findings associated with Test Anxiety are more in keeping with the suspected effects of being seen as different from one's peers. It might be that the variable of inferiority is a much more stable concept and not as easily influenced as other more "superficial" personality variables.

On this basis, it might be expected that superior pupils grouped with other pupils may experience a threat to their self-concept or personal adequacy simply on the basis of greater expectation or (more realistically) parental disapproval, or loss of peer status. Whereas low ability pupils, grouped as such, may experience an increase in ego strength and feelings of adequacy due to less peer competition and greater chances for success. Also, low-ability-level grouping may orient the pupil's parents to a better understanding of his potential and reduce parental expectation.

A second source of error might be in the scoring systems themselves. It is possible that different investigators are labeling a variable as inferiority, aggression, anxiety, etc., but may be measuring

something entirely different. Common sense is the only real evidence we have for assuming that the pupils studied by Luchins and Luchins (1948) felt an increase in feelings of inferiority upon being labeled "dumb." It is even more hypothetical to assume that the measure used in this study actually measured the same concept, level, or type of inferiority that some other researcher investigated.

The solution to this dilemma might lie in a more exact definition of "feelings of inferiority" and in the acceptance of an operational definition of the variable which any investigator could utilize.

Additional district findings

The initial hypothesis did not predict any significant differences across the CMAS (L) scale variable. However, three significant findings were found and will now be discussed:

1. Heterogeneously-grouped superior boys obtained higher CMAS (L) scale scores than superior homogeneously-grouped boys.
2. Heterogeneously-grouped average girls obtained higher CMAS (L) scale scores than homogeneously-grouped average girls.
3. Homogeneously-grouped superior girls obtained higher CMAS (L) scores than heterogeneously-grouped superior girls.

Interpretation

These findings suggest a between-district sex differential between superior pupils in both grouping systems. In District R, superior boys manifested higher (L) scores than homogeneously-grouped superior boys. The finding was opposite for girls, with homogeneously-grouped superior girls manifesting higher (L) scores than

heterogeneously-grouped superior girls. The inference is that a grouping practice may have differential effects on superior pupils which are sex determined.

Considering the (L) scale to be a measure of test-taking attitudes and a measure of the pupil's tendency to be defensive and falsify answers to universally accepted truths, the finding of a sex differential across districts and significant differences between selected groups of similar pupils in both districts strongly suggests that one method of grouping practice will not be applicable to all pupils.

Superior boys in a heterogeneous grouping situation appear to demonstrate more defensiveness than similar boys in the homogeneously-grouped situation. This finding may have relevance when we consider the finding that homogeneously-grouped superior boys obtained higher aggression scores than the superior boys in District R. The fact that District R boys are more defensive might help account for the difference. It is possible that they may be suppressing the expression of aggression. However, this implication did not hold across the variable of depression for average girls in District R who obtained higher depression scores than average girls in District A, but also obtained significantly higher CMAS (L) scores.

The findings indicated that the relationships are complex and will require a great deal more research before specific conclusions may be derived. In relation to the CMAS (L) scale, it appears that the (L) scale as a personality variable may have more discriminative ability and be a stronger tool in the study of personality structure than has

been recognized. It is noteworthy that of the six significant district findings, three were elicited by the CMAS (L) scale. Further research with the variable would probably be of value.

Ability Level Comparisons

Hypothesis No. 3 stated: "In relation to the three ability levels (superior, average, and developmental), the following significant inter-level differences will be demonstrated:

a. Superior and developmental pupils in both grouping systems will obtain higher mean scores on the personality variables of anxiety, aggression, and depression than average pupils.

b. Superior pupils in either grouping situation, as a measure of 'drive level,' will obtain higher Manifest Anxiety test mean scores; and there will be a hierarchy of Manifest Anxiety scores proceeding from superior, through average, to the developmental-- in descending order.

c. Developmental pupils will obtain higher Test Anxiety mean scores than average or superior pupils in both districts."

Anxiety, aggression, and depression

Subgroup (a) of this hypothesis was based on inferences drawn from the work of Mathias (1959) and Luchins and Luchins (1948) which suggested that pupils at the extremes of the I.Q. and achievement continuums react to the school situation and/or grouping situation with stronger feelings than the average child. On this basis, superior and developmental pupils in both districts might be expected to demonstrate

higher mean scores on the personality variables of anxiety, aggression, and depression than average pupils.

Findings

The significant findings relative to this hypothesis (summarized in Table 10) were as follows:

1. District A superior boys obtained higher mean scores on the variable of aggression than both average (5 per cent level) and developmental (1 per cent level) boys in the same district.

2. District A superior girls obtained higher mean scores on the variable of depression than developmental girls in the same district.

3. District R average girls obtained higher mean scores on the variable of depression than superior girls in the same district.

4. District R developmental boys obtained higher mean Test Anxiety and Manifest Anxiety scores than superior boys in the same district.

5. District R average boys obtained higher mean Test Anxiety scores than superior boys in the same district.

6. District R developmental girls obtained higher mean Test Anxiety scores than superior girls in the same district.

Interpretation

Whenever one works with extreme groups, it might be expected that responses to a stress situation will be more extreme than the responses of an average person in the same situation. However, the findings indicated that pupils in each district reacted differently.

Table 10. Significant ability level differences

Ability level comparisons	Personality variables					
	Aggression	Depression	Combined Inferiority	TASC	CMAS	Total CMAS (L)
District A (ability grouped)						
Boys:						
Superior vs. developmental ^a	.01 ^b					.01 ^c
Superior vs. average	.05					.05 ^c
Girls:						
Superior vs. developmental		.05				
Superior vs. average						.01
Developmental vs. average						.01
Combined boys and girls:						
Superior vs. developmental			.01			
Average vs. developmental			.01			
District R (random grouped)						
Boys:						
Average vs. superior				.05		
Developmental vs. superior				.01	.05	
Girls:						
Average vs. superior		.05				.01
Developmental vs. superior				.05		.01

^aThe first group listed obtained the highest score; higher scores are considered less desirable.

^bLevel of significance.

^cReversed relationship; latter group listed obtained higher score.

In District A (ability grouped), most of the significant level comparisons were between extreme groups--indicating that while extreme groups differ, they are not significantly different from the average groups. In those cases in which an average group differed significantly from one extreme group, the relationship did not hold true for the other extreme group--suggesting that group reaction patterns across a particular variable are selective.

The same pattern of relationships held for District R, but none of the comparisons held across districts. This suggested differential effects of the grouping system. It is interesting that superior and average pupils in the ability-grouped district tended to demonstrate more aggression and depression than developmental pupils in the same district; while in the random-grouped district, the significant difference was across the variable of anxiety, with the low-achieving pupils demonstrating significantly more Test Anxiety than the superior pupils. It appeared that in District R, low-achieving pupils were more concerned with the test situation and their performance than similar pupils in District A. It might be that homogeneous grouping practice helped reduce the concern around test anxiety in the schoolroom for the low-achieving pupils, while increasing feelings of aggression and depression in the superior pupils.

Across the variable of depression, District R average girls demonstrated significantly (5 per cent level) higher scores than District R superior girls. This is in contrast to the finding that District A superior girls obtained significantly (5 per cent level) higher depression scores than District A developmental girls. The

relationship was almost inverse and suggested differential effects of the grouping system.

It appears that in relation to the lower-ability pupils, heterogeneous grouping practice tends to increase anxiety scores on pupils of both sexes and depression scores on girls; while homogeneous grouping practice appears to have its major effect upon superior pupils, leading to higher aggression scores in boys and higher depression in girls. However, it appears that the relationship among and between the variables investigated in this study are complex, and sweeping generalities do not seem to be appropriate.

Manifest Anxiety as a measure of drive level

Subgroup (b) of the third hypothesis tested the assumption that the CMAS is a measure of drive level and postulated a hierarchy of Manifest Anxiety mean scores proceeding from superior, through average, to the developmental level--in descending order, if the assumption proved to be correct.

Findings

The one significant ability level finding was that District R developmental boys earned significantly (5 per cent level) higher CMAS scores than superior boys in the same district, which is an inverse relationship to the "drive level" hypothesis put forth by Taylor (1956). Further, the assumption that there would be a hierarchy of CMAS scores proceeding from superior, through average, to the developmental level--in descending order--was not demonstrated.

Interpretation

This hypothesis was based on the assumption put forth by Taylor (1956), which suggested that superior pupils, as a measure of "drive level," should obtain higher CMAS scores than average or developmental pupils. The fact that in the random-grouped district developmental pupils obtained higher scores than superior students raises some serious doubts as to the CMAS' ability to measure drive level. It may be that developmental boys exhibit more drive level in an attempt to meet the competition in the heterogeneous grouping system, while superior boys can keep up with little or no drive.

That the CMAS is more a measure of situational anxiety than drive level is suggested by the TASC (purported to be a measure of situational anxiety) findings. The TASC did not differentiate between districts, but the ability level differences were essentially the same as those obtained by the CMAS, in that District R average (5 per cent level) and developmental (1 per cent level) boys and developmental girls (5 per cent level) all obtained significantly higher TASC scores than District R superior pupils of either sex. This suggests that the CMAS is more a measure of situational anxiety than drive level.

The indication that District R pupils of lower scholastic standing demonstrate higher anxiety scores than superior pupils, suggests a greater situational stress upon heterogeneously-grouped pupils and, in keeping with research findings, would suggest that lower ability pupils would experience a greater decrement in learning ability than superior pupils due to the effects of increased anxiety. As District A demonstrated no significant district or level differences in CMAS or TASC

scores, one might assume that the practice of homogeneously grouping pupils on an ability level may have the effect of reducing situational anxiety and of contributing to the learning efficiency of the pupil. This would be a valuable area of further research.

The hypothesis of a hierarchy of CMAS scores as indicative of drive level was not supported, and in those cases where the difference was significant, the relationship was inverse to that postulated in the drive theory. In District A, wherein the differences would be expected to be more clear-cut because of the grouping system, the predicted tendency was non-existent. Considering Test Anxiety alone, it appeared that pupils in the heterogeneous setting were more threatened (anxious) by their relationship with superior students than in the homogeneous grouping system. However, the lack of any significant district differences suggests that the difference is quite small. A possible explanation is in keeping with the findings of Sarnoff et al. (1956) who found that as pupils became accustomed to the situation, the level of anxiety is reduced. As the pupils in District A had been associated with grouping practices for over two years, it might be that the pupils accepted roles and were not threatened by the grouping practice.

Test Anxiety

Hypothesis No. 3 (c) stated that developmental pupils will obtain higher Test Anxiety mean scores than the average or superior pupils in both districts.

Findings

This hypothesis held true only in the heterogeneously-grouped district (District R), and only on the relationship between developmental (1 per cent level) and average (5 per cent level) boys and developmental (5 per cent level) girls and superior pupils of both sexes.

Interpretation

The basis for this assumption was that developmental pupils in either grouping system would be threatened by any test situation (on the basis of poor performance) and demonstrate higher TASC scores. The findings suggested that, on a level basis, pupils in homogeneously-grouped classes are less threatened by test situations than heterogeneously-grouped students. The reasons for the difference may be many, but it seems reasonable that the tests constructed by teachers for homogeneously-grouped developmental pupils might be geared more to their level of ability and offer these pupils more opportunities for success than heterogeneously-grouped pupils who must compete on tests geared at least to the ability of average students. Also, homogeneously-grouped students might feel more accepting of a poor test performance on the basis that they have been accepted as slow learners. It is likely that in the heterogeneously-grouped classroom low ability students are not so accepted. Again, the absence of district differences suggested that the level difference is small.

Additional level findings

Level findings not hypothesized were found in two different areas.

Feelings of inferiority

Level differences across the variable feelings of inferiority were not hypothesized; however, significant differences in this variable were demonstrated.

Findings. Combining sexes by level across the personality variable feelings of inferiority (procedure explained in Hypothesis No. 2, subgroup (c), and in the Methods and Procedures chapter) produced the following findings: District A superior and average pupils obtained significantly higher (1 per cent level) mean scores on the variable of feelings of inferiority than developmental pupils in the same district.

Interpretation. These findings suggest that superior and average boys and girls demonstrate more feelings of inferiority in homogeneous grouping situations than pupils of low ability in the same grouping arrangement. It has already been demonstrated that District A superior boys earned higher mean aggression scores than District A developmental boys, and the relationship held for District A superior and developmental girls around the variable of depression. It is generally accepted that in our society boys demonstrate more aggression than girls, and that girls obtain higher depression scores than boys. It might be that superior girls in the homogeneous grouping situation

react to feelings of inferiority with depression, while homogeneously-grouped superior boys react to feelings of inferiority by demonstrating more aggression. The finding suggests that homogeneous grouping practices may be producing inter-level differences which would not appear to be constructive and is having its primary effect upon the superior pupil. It may be that the superior pupils in the homogeneous grouping system are associated with a change from an easy competitive situation in their first years of schooling (heterogeneously grouped) to a more competitive difficult situation after grouping started. This could have a threatening effect upon the superior pupil as he now has to compete with more intelligent pupils. Assuming that aggression may be a compensation for feelings of inferiority, and assuming that feelings of inferiority are detrimental to the progress and development of the pupil, one might seriously question the practice of homogeneously grouping superior pupils. The modern-day emphasis upon grouping superior students in order to offer them the "benefits" of accelerated training may be having a deleterious effect upon the superior pupil. It may be that homogeneous grouping is of more value to the low-ability pupil and of less value to the superior pupil.

CMAS (L) scale

The CMAS (L) scale was included in the present test battery as a screening device intended to identify those pupils who were inclined to adopt a negative response set and falsify or distort their test responses. Available research indicates that lower ability students tend to obtain higher CMAS (L) scale scores and that higher (L) score

pupils are inclined to be less popular and more often rejected by their peers (Trent, 1957). As these studies were not considered conclusive, and as the (L) scale was included primarily as a screening measure, hypotheses pertaining to this variable were not developed. However, significant differences across the (L) scale variable did occur.

Findings. The significant CMAS (L) scale findings were as follows:

1. District A developmental (1 per cent level) and average (5 per cent level) boys obtained higher mean (L) scale scores than superior boys in the same district.

2. District A superior (1 per cent level) and developmental (1 per cent level) girls obtained higher mean (L) scale scores than average girls in the same district.

3. District R average (1 per cent level) and developmental (1 per cent level) girls obtained higher mean (L) scale scores than superior girls in the same district.

Interpretation. The findings indicate that District A average (5 per cent level) and developmental (1 per cent level) boys demonstrated a greater tendency towards falsification than superior students in the same district. In contrast, District A superior (1 per cent level) and developmental (1 per cent level) girls demonstrated higher (L) scale scores than average girls in the same district. This suggests that lower level District A boys tend to falsify more than superior boys in the same district. This is in keeping with research findings which

suggest a negative relationship between high (L) scale scores and academic achievement. However, the relationship did not hold true for District A girls, in that both superior and developmental girls obtained higher (L) scale scores than average District A girls. The suggestion is that District A superior boys have less need to defend against manifesting anxiety than developmental boys in the same district; whereas District A girls, on either end of the grouping continuum, feel a greater need to deny anxiety than average girls in the same district.

In the District R heterogeneously-grouped situation, the findings were more in keeping with research suggestions, in that both average and developmental girls demonstrated significantly (1 per cent level) higher (L) scale scores than superior girls in the same district. There were no significant level differences between the boys.

Sex Differences

Hypothesis No. 4 stated: "There will be no significant sex differences, except the following:

- a. Boys at all levels will obtain higher mean scores on the personality variable of aggression than girls.
- b. Girls at all levels will obtain higher mean scores on the personality variable of depression than boys.
- c. Girls at all levels will obtain higher Manifest Anxiety test scores than boys.
- d. Girls at all levels will obtain higher mean Test Anxiety scores than boys.
- e. Girls at all levels will obtain higher mean (L) lie scores than boys on the Manifest Anxiety (L) scale."

Table 11 presents a summary of the significant sex differences found in this study.

Aggression

Subgroup (a) of this hypothesis stated that boys at all levels will obtain higher mean scores on the personality variable of aggression than girls.

Findings

The only significant sex difference finding across the variable of aggression was that homogeneously-grouped superior boys demonstrated more aggression than homogeneously-grouped superior girls.

Interpretation

This finding was in keeping with research findings, except that it should have extended across all levels and districts. The fact that it did not, suggests something inherent in the homogeneous grouping system. Feelings of inferiority have been suggested as the reason for this difference.

Depression

Subgroup (b) of this hypothesis stated that girls at all levels would manifest higher mean scores on the personality variable of depression than boys.

Table 11. Significant sex differences

Sex comparisons	Personality variables					
	Aggression	Depression	Combined Inferiority	TASC	CMAS	Total CMAS (L)
District A						
<u>Superior level:</u>						
Boys vs. girls ^a	.05 ^b	.05 ^c		.05 ^c	.05 ^c	.01 ^c
<u>Average level:</u>						
Girls vs. boys				.01	.05	
<u>Developmental level:</u>						
Girls vs. boys						.05
District R						
<u>Superior level:</u>						
Girls vs. boys				.05	.01	
<u>Average level:</u>						
Girls vs. boys		.01			.01	
<u>Developmental level:</u>						
Girls vs. boys						.05

^aThe first group listed obtained the highest score; higher scores are considered less desirable.

^bLevel of significance.

^cThe relationship is reversed; the latter group listed obtained higher scores than the first.

Findings

The findings were in the predicted directions, but were significant only for selected groups. In District R average girls demonstrated significantly (1 per cent level) higher depression scores than average boys in the same district, but there were no other significant sex differences.

In District A, superior girls demonstrated significantly higher (5 per cent level) depression scores than superior boys in that same district.

Interpretation

There is no apparent reason why average girls in District R, as a selected group, should obtain higher scores on the variable of depression than average boys in the same district. It is noteworthy that District R average girls also demonstrated higher Manifest Anxiety scores than average boys in the same district, and it is possible that there is something inherent in the grouping system that contributes to this condition; however, the cause is not discernable from the data derived from this study.

The finding that District A superior girls demonstrated more depression than superior boys in that district is in keeping with the general findings that girls customarily demonstrate more depression than boys. However, a more complex condition is suggested by the fact that superior girls also obtained higher Test Anxiety, Manifest Anxiety, and CMAS (L) scores than superior boys in the same district. Also, one would question why only the superior girls in District A

demonstrated the predicted sex difference. It seems apparent that the superior girls in District A are affected by the grouping situation and react to it in a manner that would seem to be detrimental to their best performance. Increased feelings of inferiority have been suggested as a contributing factor, but the inference is not conclusive.

The fact that few significant differences were obtained suggests that the general effects of grouping practice is in most cases constant, with the majority of pupils relating to either particular system in a fairly uniform manner. The absence of clear-cut district differences suggests that except in special cases the specific effect of a particular grouping practice is not significant in terms of the variables studied.

Manifest Anxiety scores

Subgroup (c) of Hypothesis No. 4 stated that girls at all levels will obtain higher mean Manifest Anxiety test scores than boys.

Findings

This hypothesis held for both superior and average girls in both districts--but did not hold for the developmental girls in either district.

Interpretation

Research has suggested that girls in our culture customarily demonstrate higher Manifest Anxiety test scores than boys, primarily because our society condones this type of expression for girls, but inhibits such expression in boys by considering such behavioral manifestations as unmanly. The present findings indicate that in this

sample of sixth-grade pupils, developmental boys in both districts demonstrated as much Manifest Anxiety as girls. A possible reason for this might be that low-ability boys may have fewer methods and means of demonstrating adequacy simply because of limited ability, and any situation which is likely to evoke a failure could precipitate greater anxiety. This supposition is given impetus when one remembers that District R developmental boys obtained higher CMAS ("drive level") scores than District R superior boys, which may suggest the importance developmental pupils place around achievement--and, inversely, the "fear" around failure.

However, a comparison of the non-significant mean scores earned by developmental pupils in both districts demonstrated a trend toward sex differences that were in keeping with the hypothesis. Table 12 presents these comparisons.

Table 12. Non-significant mean CMAS score sex differences for developmental pupils in both districts

Sex differences	Mean CMAS scores	
	Girls	Boys
Developmental girls vs. developmental boys		
District R	15.12	14.07
District A	17.54	13.27

While not significant, the trend suggests that girls at all levels earn higher Manifest Anxiety test scores than boys, regardless of the grouping system--which suggests that the expression of anxiety

is at least in part a culturally-determined phenomenon.

Test Anxiety

Subgroup (d) of the fourth hypothesis stated that girls at all levels will obtain higher mean Test Anxiety scores than boys.

Findings

The significant sex-determined Test Anxiety findings are as follows:

1. District A superior girls obtained higher Test Anxiety scores than superior boys in the same district.
2. District A average girls obtained higher Test Anxiety scores than average boys in the same district.
3. District R superior girls obtained higher Test Anxiety scores than superior boys in the same district.

Whenever a significant difference was indicated, it was in the predicted direction: girls at all levels obtained higher mean Test Anxiety scores than boys, but the difference was not significant at all levels.

However, the results were not consistent with the CMAS results, in that, in general, superior and average girls in both districts demonstrated more anxiety than boys. The developmental group in either district did not demonstrate a significant sex difference; however, a comparison of the non-significant mean scores earned by developmental pupils on the TASC and CMAS tests demonstrated a trend toward sex differences which were in keeping with the hypothesis.

Table 13 presents these comparisons.

Table 13. Non-significant mean TASC and CMAS score sex differences for developmental pupils in both districts

Sex differences	Mean Scores			
	TASC		CMAS	
	Girls	Boys	Girls	Boys
Developmental ability level				
District R	14.28	10.26	15.12	14.07
District A	14.08	10.40	17.54	13.27

Interpretation

These findings demonstrate that while not significant, developmental girls in both districts earned consistently higher anxiety scores than developmental boys. Considering a level difference, for the purpose of clarification, the TASC findings roughly parallel the CMAS findings in that developmental boys and girls in District R demonstrated significantly higher TASC scores than superior District R pupils of the same sex in that district. This suggests that being a developmental student in either grouping system has the effect of raising the anxiety level of both boys and girls in the heterogeneous setting and increasing the anxiety level of homogeneously-grouped developmental boys. These findings could also reflect a reduction of anxiety scores in developmental girls. However, a trend analysis of non-significant relationships supports the hypothesis that girls in general demonstrate more anxiety than boys at any level.

(L) Lie scale

Subgroup (e) of Hypothesis No. 4 stated that girls at all levels will obtain higher mean (L) scores than boys on the CMAS (L) scale.

Findings

The significant findings were as follows:

1. District A superior girls obtained higher mean (L) scores than superior boys in the same district.
2. District A developmental girls obtained higher mean (L) scores than developmental boys in the same district.
3. District R developmental girls obtained higher mean (L) scores than developmental boys in the same district.

Interpretation

Research evidence indicates that girls customarily obtain higher CMAS (L) scores than boys. The present study indicated that in relation to sex differences, District A superior (1 per cent level) and developmental (5 per cent level) girls demonstrated significantly higher (L) scores than boys at the same levels. The average group demonstrated no sex differences in either district.

In District R, developmental girls demonstrated higher (L) scores than developmental boys (5 per cent level), but no other significant differences were indicated.

This suggests that superior and developmental girls in a homogeneous grouping situation tend to falsify their responses to the CMAS more than boys in the same grouping situation. The tendency

towards falsification suggests a response set towards defensiveness and would suggest that the girls at these levels are more concerned with their performance than boys.

SUMMARY, FINDINGS, CONCLUSIONS, AND RECOMMENDATIONS

Summary

The primary purpose of this study was to investigate the differential effects of heterogeneous versus homogeneous grouping practices on the personality structure of a selected sample of sixth-grade public school children. The sample was drawn from two adjacent school districts in which District A groups homogeneously, according to ability, and District R utilizes a random ability-grouping method. The personality variables studied were anxiety, aggression, depression, and feelings of inferiority.

A second purpose of this research was to test the reliability of a partially objectified, inter-rater type rating scale method of scoring Thematic-Apperception-Test-type stories. The developed rating scale was used to evaluate the personality variables of aggression, depression, and feelings of inferiority. Anxiety was evaluated by two objective-type test questionnaires.

The research design allowed for the investigation of the differential effects of three conditions of grouping practice (district, level, and sex) on the measurement of six selected personality indices.

Development of scoring method

A partially-objectified, inter-rater type rating scale method of scoring TAT-type stories that would (a) take advantage of cumulative data elicited from TAT-type stories; (b) allow for the simultaneous

scoring of more than one personality variable; (c) not be dependent upon concrete response, word counts, or other structural approaches; (d) allow the use of clinical insight, conclusions, and interpretations; (e) elicit no negative scores; and (d) allow all impressions to be additive was developed.

Two raters, using the developed method of scoring, scored 338 seven-story TAT-type protocols with sufficient inter-rater reliability to permit group comparison across the personality variables of aggression, depression, and feelings of inferiority.

Methods and procedures

A sample of 338 sixth-grade pupils was randomly selected from three school populations (superior, average, and below-average) in two adjacent school districts. School District A employed homogeneous (ability) grouping, and District R employed a heterogeneous (random) grouping system.

Pupils were administered a battery of three tests which were used in the investigation of the four personality variables: aggression, depression, feelings of inferiority, and anxiety.

A modified 7-picture version of the TAT was presented by means of an opaque projector and pupils were asked to write 7-minute stories to each stimulus picture. The stories were scored according to the developed scoring method, and mean group scores were determined.

The CMAS and TASC tests were scored according to the standard method, and mean group scores were determined.

The means of 12 different conditions (groups) were compared by a simple analysis of variance technique and the resulting significant variance ratios, F , were tested for separate differences by the standard t test.

Findings

The following results were obtained from the investigation:

Scoring of TAT-type stories

In regard to Hypothesis No. 1, which stated that Thematic-Apperception-Test-type stories can be scored by a five-point rating scale with sufficient accuracy to permit inter- and intra-group comparisons on the personality variables of aggression, depression, and feelings of inferiority, the findings were positive. Two raters independently scored 338 seven-story TAT-type protocols obtained from sixth-grade pupils, according to a previously-developed scoring method, with sufficient accuracy to permit inter-group comparisons across the personality variables of aggression, depression, and feelings of inferiority. The derived inter-rater correlations (corrected by the Spearman-Brown Prophecy Formula) were as follows: aggression, .91; depression, .70; and inferiority, .73.

District comparisons

Hypothesis No. 2 stated that in relation to the selected personality variables (anxiety, aggression, depression, and feelings of inferiority) there will be no significant differences between similar

pupils in the two districts, except for the following:

a. Superior and developmental pupils in the heterogeneously-grouped classroom will obtain higher mean scores across the variables of aggression and depression than similar pupils in homogeneously-grouped classroom as an effect of frustrations imposed by the restrictions of a school situation that is geared to the needs of the average pupil.

b. Homogeneously-grouped superior pupils, as a result of intra-group competition and fears associated with losing social status if not successful, will obtain higher Manifest and Test Anxiety mean scores than heterogeneously-grouped superior pupils.

c. Homogeneously-grouped developmental pupils will, as a result of being classified as "dumb," obtain higher mean scores on the personality variable "feelings of inferiority" than similar heterogeneously-grouped pupils.

The significant district findings were as follows:

1. District A superior boys demonstrated more aggression than District R superior boys.

2. District R average girls obtained higher depression scores than average girls in District A.

3. District R developmental girls demonstrated more depression than developmental girls in District A.

4. District R superior boys obtained higher CMAS (L) scores than District A superior boys.

5. District A superior girls obtained higher CMAS (L) scores than District R superior girls.

6. District R average girls obtained higher CMAS (L) scores than District A average girls.

Ability level comparisons

Hypothesis No. 3 stated that in relation to the three ability levels (superior, average, and developmental) the following significant inter-level differences will be demonstrated:

a. Superior and developmental pupils in both grouping systems will obtain higher mean scores on the personality variables of anxiety, aggression, and depression than average pupils.

b. Superior pupils in either grouping situation, as a measure of "drive level," will obtain higher Manifest Anxiety test mean scores; and there will be a hierarchy of Manifest Anxiety scores proceeding from superior, through average, to the developmental-- in descending order.

c. Developmental pupils will obtain higher Test Anxiety mean scores than average or superior pupils in both districts.

The significant ability level findings were as follows:

1. Similar pupils in both districts reacted differently.

2. District A superior boys obtained higher scores on the variables of aggression and inferiority than District A developmental boys.

3. District A developmental boys obtained higher CMAS (L) scores than District A superior boys.

4. District A superior boys obtained higher aggression scores than District A average boys.

5. District A average boys obtained higher CMAS (L) scores than District A superior boys.
6. District R average girls obtained higher depression and CMAS (L) scores than District R superior girls.
7. District A superior girls obtained higher depression and feelings of inferiority scores than District A developmental girls.
8. District A average boys obtained higher feelings of inferiority scores than District A developmental boys.
9. District A superior and average pupils both obtained higher combined feelings of inferiority scores than District A developmental pupils.
10. District R developmental boys obtained higher TASC and CMAS scores than District R superior boys.
11. District R average boys obtained higher TASC scores than District R superior boys.
12. District R developmental girls obtained higher TASC and CMAS (L) scores than District R superior girls.
13. District A superior girls obtained higher CMAS (L) scores than District A average girls.

Sex differences

Hypothesis No. 4 stated that there will be no significant sex differences, except the following:

- a. Boys at all levels will obtain higher mean scores on the personality variable of aggression than girls.
- b. Girls at all levels will obtain higher mean scores on the

personality variable of depression than boys.

c. Girls at all levels will obtain higher Manifest Anxiety test scores than boys.

d. Girls at all levels will obtain higher mean Test Anxiety scores than boys.

e. Girls at all levels will obtain higher mean (L) lie scores than boys on the Manifest Anxiety (L) scale.

The significant sex difference findings were as follows:

1. District A superior boys obtained higher aggression scores than District A superior girls.

2. District A superior girls obtained higher depression, TASC, CMAS, and CMAS (L) scores than District A superior boys.

3. District R average girls obtained higher depression and CMAS scores than District R average boys.

4. District A average girls obtained higher TASC and CMAS scores than District A average boys.

5. District R superior girls obtained higher TASC and CMAS scores than District R superior boys.

6. District R developmental girls obtained higher CMAS (L) scores than District R developmental boys.

7. District A developmental girls obtained higher CMAS (L) scores than District A developmental boys.

Conclusions

On the basis of the findings, it was concluded that--

1. TAT-type stories can be scored by a five-point, partially objectified, inter-rater type rating scale with sufficient accuracy to permit inter- and intra-group comparisons on the personality variables of aggression, depression, and feelings of inferiority.

2. Insofar as the findings of this study are representative, the effects of a grouping practice do not appear to be global. The question as to whether one should or should not group sixth-grade pupils appears to reduce itself to the level of asking which level and which pupils should be grouped and which should not be grouped, rather than to the level of which system is best.

3. Superior boys in the homogeneously-grouped situation demonstrate higher aggression scores than both average and developmental pupils in the same district and superior boys in heterogeneously-grouped situations.

4. Heterogeneously-grouped average and developmental girls demonstrate higher depression scores than similar girls in the homogeneously-grouped situation.

5. In relation to the personality variables of anxiety and feelings of inferiority, there are no significant differences between similar pupils in both districts.

6. Both superior and average boys and girls in the homogeneously-grouped situations demonstrate more feelings of inferiority than developmental pupils in the same grouping situation.

7. Heterogeneous grouping may have a significant effect upon the lower ability pupils by increasing their anxiety, while homogeneous grouping may have its major effect upon the superior pupils by leading to more aggression in boys and more depression in girls.

8. The CMAS appears to be more a measure of situational anxiety than a measure of "drive level" and appears to be measuring a variable similar to that measured by the TASC.

9. In general, girls in either grouping situation will obtain higher TASC and CMAS test scores than similar boys at any ability level.

10. As measured by the CMAS (L) scale, homogeneously-grouped superior and developmental girls demonstrate a stronger response set towards falsifying their responses to the CMAS items than homogeneously-grouped boys of similar ability level.

11. Heterogeneously-grouped superior boys obtain higher CMAS (L) scores than homogeneously-grouped superior boys. However, the relationship is reversed for superior girls, with homogeneously-grouped superior girls obtaining higher CMAS (L) scores than superior girls in the heterogeneously-grouped situation.

12. Homogeneously-grouped average and developmental boys obtain higher CMAS (L) scores than superior boys in the same district. In contrast, homogeneously-grouped superior and developmental girls obtain higher CMAS (L) scores than average girls in the same district.

Recommendations

On the basis of the findings, the following recommendations for future research are set forth:

1. The relationships between the sex of the pupil, the level of academic achievement, and the method of grouping practice and its affect upon the personality development of the pupil should be investigated more completely. The present findings suggest that the effects of a grouping practice upon a particular pupil are dependent upon or a function of a pupil's sex and the level of his academic achievement, with different combinations being related to different patterns of personality adjustments.

2. The value and the meaning of the CMAS (L) score as a critical measure of individual and group personality differences should be further investigated. It was noted that of the 35 significant findings obtained in the study, 12 (34.2 per cent) were on the CMAS (L) scale. It might be that the CMAS (L) scale has more discriminative value than has previously been acknowledged.

3. The establishment of operational definitions for complex psychological terms that could be used by all investigators would reduce much of the ambiguity in many research studies.

BIBLIOGRAPHY

- Abt, L. E., and L. Bellak. 1950. Projective psychology. New York: Alfred A. Knopp.
- Atkinson, John W. 1958. Motive in fantasy, action, and society. Princeton, N. J.: D. Van Nostrand Co., Inc.
- Auld, Frank, Jr., L. D. Eron, and J. Laffal. 1955. Application of Guttman's scaling method to the TAT. Educational and Psychological Measurement, XV: 422-35.
- Barthelmess, H. M., and P. A. Boyer. 1932-33. An evaluation of ability grouping. Journal of Educational Research, XXVI:284-94.
- Bedoian, V. H. 1954. Social acceptability and social rejection of the underage, at age, and over age pupils in the sixth grade. Journal of Educational Research, XLVII:513-20.
- Bendig, A. W. 1949. The development of a short form of the Manifest Anxiety Scale. Journal of Consulting Psychology, XX:384.
- Bettelheim, B. 1958. Segregation: New style. School Review, LXVI:251-72.
- Castaneda, Alfred, Boyd R. McCandless, and David S. Palermo. 1956a. The children's form of the Manifest Anxiety Scale. Child Development, XXVII:317-26.
- Castaneda, Alfred, Boyd R. McCandless, and David S. Palermo. 1956b. Complex learning and performance as a function of anxiety in children and task difficulty. Child Development, XXVII:327-32.
- Cook, R. R. 1924. A study of the results of homogeneous grouping of abilities in high school classes. National Society for the Study of Education, XXIII:302-12.
- Coombs, Arthur W. 1947. A comparative study of motivation as revealed in Thematic Apperception stories and autobiography. Journal of Clinical Psychology, III:65-75.
- Cowen, E. L., et al. 1957. The correlates of manifest anxiety in perceptual reactivity, rigidity, and self concepts. Journal of Consulting Psychology, XXI:405-11.
- Dana, Richard H. 1955. Clinical diagnosis and objective TAT scoring. Journal of Abnormal and Social Psychology, L:19-24.

- Dana, Richard H. 1956. An application of objective TAT scoring. *Journal of Projective Psychology*, XX:159-63.
- Davids, Anthony, and Charles W. Eriksen. 1958. The relation of manifest anxiety to association productivity and intellectual attainment. *Journal of Consulting Psychology*, XIX:219-22.
- Davids, Anthony, et al. 1955. Projection, self evaluation, and clinical evaluation of aggression. *Journal of Consulting Psychology*, XIX:437-40.
- Elliott, Jaques C. 1945. The clinical use of the TAT with soldiers. *Journal of Abnormal and Social Psychology*, XL:363-75.
- Engle, T. L. 1938. A study of the effects of school acceleration upon the personality and social adjustments of high school and university students. *Journal of Educational Psychology*, XII:523-39.
- Eron, L. D. 1950. A normative study of the TAT. *Psychological Monograph*, IX:64.
- Eron, L. D., and A. M. Ritter. 1951. A comparison of two methods of administration of the Thematic Apperception Test. *Journal of Consulting Psychology*, XV:55-61.
- Eron, L. D., Dorothy Terry, and R. Callahan. 1950. Use of rating scales for emotional tone of TAT stories. *Journal of Consulting Psychology*, XIV:473-78.
- Essman, W. B. 1957. Manifest anxiety and test performance. *Perception and Motor Skills*, VII:214.
- Eysenck, H. J. 1961. *Handbook of abnormal psychology*. New York: Basic Books, Inc.
- Fine, Reuben. 1955a. Manual for a scoring scheme for verbal projective techniques (TAT, MAPS stories and the like). *Journal of Projective Techniques*, XIX:310-16.
- Fine, Reuben. 1955b. A scoring scheme for the TAT and other verbal projective techniques. *Journal of Projective Techniques*, XIX:306-09.
- French, Elizabeth G., and Irene Chadwick. 1956. Some characteristics of affiliation motivation. *Journal of Abnormal and Social Psychology*, LII:296-300.
- Fridman, Ira. 1957. Objectifying the subjective: A methodological approach to the TAT. *Journal of Projective Techniques*, XXI:243-47.
- Friedman, G. A. 1950. A cross-cultural study of the relationship between independence training and n Achievement as revealed by mythology. Unpublished honors thesis, Harvard University, Cambridge, Mass.

- Gallagher, James J. 1958. Peer acceptance of highly gifted children in elementary schools. *The Elementary School Journal*, May, pp. 465-70.
- Garrett, Henry E. 1959. *Statistics in psychology and education*. New York: Longmans, Green and Co.
- Goodlad, John I. 1954. Some effects of promotion and non-promotion upon the social and personal adjustment of children. *Journal of Experimental Education*, XI:301-28.
- Goodstein, Leonard. 1954. Interrelationships among several measures of anxiety and hostility. *Journal of Consulting Psychology*, XVIII:35-39.
- Gowan, May Seago. 1955. Why homogeneous grouping? *California Journal of Secondary Education*, XXX:22-28.
- Gronlund, Norman E., and Loren Anderson. 1957. Personality characteristics of socially accepted, socially neglected, and socially rejected junior high school pupils. *Education, Administration, and Supervision*, XLIII:329-38.
- Guttman, Louis. 1944. A basis for scaling quantitative data. *American Sociological Review*, IX:139-50.
- Hafner, A. J. and A. M. Kaplan. 1959. An item analysis of the Children's Manifest Anxiety Scale. *Child Development*, XXX:48-88.
- Harrison, Ross, and Julian B. Rotter. 1945. A note on the reliability of the TAT. *Journal of Abnormal and Social Psychology*, XL:97-99.
- Henry, William E., and Jane Farley. 1959. A study in validation of the TAT. *Journal of Projective Techniques*, XXIII:273-77.
- Hokanson, Jack E., and Jesse E. Gordon. 1958. The expression and inhibition of hostility in imaginative and overt behavior, LVII: 327-33.
- Hunt, W. 1942. Ability grouping. *Education, Administration, and Supervision*, XXVIII:591-99.
- Johns, LaVerd. 1960. The relationship of the achievement motives as projected in the TAT stories to children's school achievement. Unpublished Ed.D. dissertation, Utah State University, Logan, Utah.
- Kagan, Jerome. 1956. The measurement of overt aggression from fantasy. *Journal of Abnormal and Social Psychology*, LII:390-93.
- Kurtz, John J., and E. J. Swenson. 1951. Factors related to over-achievement and under-achievement in school. *School Review*, LIX:472-80.

- Lesser, Gerald S. 1958a. Application of Guttman's scaling method to aggressive fantasy in children. *Educational and Psychological Measurement*, XVIII:543-52.
- Lesser, Gerald S. 1958b. Conflict analysis of fantasy aggression. *Journal of Personality*. (In press references obtained in *Educational and Psychological Measurement*, XVIII:29.)
- Levy, Nissim. 1958. A short form of the Children's Manifest Anxiety Scale. *Child Development*, XXIX:1953-54.
- Lichtenberg, Philip. 1957. A definition and analysis of depression. *American Medical Association Archives of Neurology and Psychiatry*, LXX:519-27.
- Lindzey, Gardner, and Arthur S. Newberg. 1954. Thematic Apperception Test: A tentative appraisal of some signs of anxiety. *Journal of Consulting Psychology*, XVIII:389-95.
- Lindzey, Gardner, and Charlotte Tejessy. 1956. Thematic Apperception Test: Indices of aggression in relation to measures of overt and covert behavior. *American Journal of Orthopsychiatry*, XXVI:567-76.
- Little, Kenneth B., and Edwin S. Schneidman. 1955. The validity of thematic projective technique interpretations. *Journal of Personality*, XXIII:285-94.
- Luchins, A. S., and E. H. Luchins. 1948. Children's attitudes toward homogeneous grouping. *Pedagogical Seminary*, LXXII:3-9.
- Mandler, George, Gardner Lindzey, and R. G. Crouch. 1957. Thematic Apperception Text: Indices of anxiety in relation to test anxiety. *Educational and Psychological Measurement*, XVII:466-74.
- Mandler, George, and Seymour B. Sarason. 1952a. The effect of prior experience and subjective experience on the evocation of test anxiety. *Journal of Personality*, XXI:336-41.
- Mandler, George, and Seymour B. Sarason. 1952b. A study of anxiety and learning. *Journal of Abnormal and Social Psychology*, XLVII:166-73.
- Mann, M. 1960. What does ability grouping do to the self-concept. *Childhood Education*, XXXVI:357-60.
- Marsh, D. E. 1955. An interesting experience in grouping. *California Journal of Secondary Education*, XXX:49-59.
- Matarazzo, J. D., et al. 1954. The relationship between anxiety level and several measures of intelligence. *Journal of Consulting Psychology*, XVIII:201-05.

- Mathias, Rudolf. 1959. Methodology and results of personality appraisal with Rorschach, TAT and figure drawing. Paper presented at American Psychological Association Convention in Cincinnati, Ohio.
- Mayzner, Mark S., Eugene Lersen, and M. E. Tresselt. 1955. The Taylor Manifest Anxiety Scale and intelligence. *Journal of Consulting Psychology*, XIX:401-03.
- McArthur, Charles. 1953. The effects of need achievement on the content of TAT stories. *Journal of Abnormal and Social Psychology*, XLIII:532-36.
- McCandless, Boyd R., and Alfred Castaneda. 1956. Anxiety in children, school achievement, and intelligence. *Child Development*, XXVII: 379-80.
- McCandless, Boyd R., Alfred Castaneda, and David S. Palermo. 1956. Anxiety in children and social status. *Child Development*, XXVII: 385-90.
- McClelland, David C. 1955. Measuring Motivation in phantasy. *Studies in Motivation*. New York: Appleton-Century-Crofts, Inc.
- Muggenthaler, F. 1955. The sociological status of slow learning pupils in an elementary grade. *Prax Kinder-psychologie Kinder-psychiatrie*, IV:226-30.
- Murray, H. A. 1943. Thematic Apperception Test manual. Cambridge, Mass.: Harvard University Printing Office.
- Murray, H. A. 1955. Types of human needs. *Studies in Motivation*. David C. McClelland (ed.). New York: Appleton-Century-Crofts, Inc.
- Mussen, P. H. 1960. Handbook of research methods in child development. New York: John Wiley and Sons, Inc.
- Mussen, P. H., and Mary C. Jones. 1957. Self conception, motivation, and interpersonal attitudes of late and early maturing boys. *Child Development*, XXVIII:243-56.
- Mussen, P. H., and H. Kelly Naylor. 1954. The relationship between overt and fantasy aggression, XLIX:235-40.
- Mussen, P. H., and Alvin Scodel. 1955. The effects of sexual stimulation under various conditions on the TAT sexual responsiveness. *Journal of Consulting Psychology*, XIX:90.
- Palermo, David S., Alfred Castaneda, and Boyd R. McCandless. 1956. The relationship of anxiety in children to performance in a complex learning task. *Child Development*, XXVII:333-37.

- Pittluck, Patricia. 1950. The relationship between aggressive fantasy and overt behavior. Unpublished Ph.D. dissertation, Yale University, New Haven, Conn.
- Purcell, Kenneth. 1956. The TAT and antisocial behavior. *Journal of Consulting Psychology*, XX:449-56.
- Rudd, W. G. A. 1958. The psychological effects of streaming by attainment. *British Journal of Educational Psychology*, XXVIII:47-60.
- Sarason, Irwin G. 1957. The effect of anxiety and two kinds of failure on serial learning. *Journal of Personality*, XXV:383-92.
- Sarason, Seymour B. 1961. Anxiety and learning. Human variability and learning. Washington 6, D. C.: Association for Supervision and Curriculum Development, Dept. of National Education Assoc.
- Sarason, Seymour B. and Edward M. Gordon. 1953. The test anxiety questionnaire: Scoring norms. *Journal of Abnormal and Social Psychology*, XLVIII:447-48.
- Sarason, Seymour B., et al. 1958a. Classroom observations of high and low anxious children. *Child Development*, XXIX:287-95.
- Sarason, Seymour B., et al. 1958b. Rorschach behavior and performance of high and low anxious children. *Child Development*, XXIX:277-85.
- Sarason, Seymour B., et al. 1958c. A test anxiety scale for children. *Child Development*, XXIX:105-113.
- Sarason, Seymour B., et al. 1960. Anxiety in elementary school children. New York: John Wiley and Sons, Inc.
- Sarnoff, Irving, et al. 1956. Test anxiety and the "Eleven Plus" examinations. *British Journal of Educational Psychology*, XXIX:9-16.
- Sarnoff, Irving, et al. 1958. A cross cultural study of anxiety among American and English school children. *Journal of Educational Psychology*, XLIX:129-36.
- Schultz, R. E., and Allen D. Calvin. 1955. A failure to replicate the finding of a negative correlation between manifest anxiety and A.C.E. scores. *Journal of Consulting Psychology*, XIX:223-24.
- Scodel, A., and M. E. Lipetz. 1957. TAT hostility and psychopathology. *Journal of Projective Techniques*, XXI:161-65.
- Silverman, Robert E., and Bernard Blitz. 1956. Learning and two kinds of anxiety. *Journal of Abnormal and Social Psychology*, LII:301-03.

- Sinick, Daniel J. 1956. Two anxiety scales correlated and examined for sex differences. *Journal of Clinical Psychology*, XII:394-95.
- Stone, Harold. 1957. The TAT aggressive content scale. *Projective Techniques*, XX:445-52.
- Taylor, Janet A. 1953. A personality scale of manifest anxiety, also called the biographical inventory. *Abnormal Social Psychology*, XLVIII:285-90.
- Taylor, Janet A. 1956. Drive theory and manifest anxiety. *Psychological Bulletin*, LIII:303-20.
- Taylor, Janet A. 1958. The effects of anxiety level and psychological stress on verbal learning. *Journal of Abnormal and Social Psychology*, LVII:55-60.
- Tonsor, C. A. 1953. Must we segregate? *National Association of Secondary School Principals Bulletin*, XXXVII:75-77.
- Trent, Richard D. 1957. The relationship of anxiety to popularity and rejection among institutionalized delinquent boys. *Child Development*, XXVIII:379-84.
- Waite, R. R., et al. 1958. A study of anxiety and learning in children. *Journal of Abnormal Social Psychology*, LVII:267-70.
- Willman, Phyllis. 1945. The use of multiple-choice Rorschach as a differential diagnostic tool. *Journal of Clinical Psychology*, I:281-87.
- Wirt, Robert D., and William E. Broen, Jr. 1956. The relation of the Children's Manifest Anxiety Scale to the concept of anxiety as used in the clinic. *Journal of Consulting Psychology*, XX:482.
- Zweibelson, I. 1956. Test anxiety and intelligence test performance. *Journal of Consulting Psychology*, XX:479-81.

A P P E N D I X

TEST ADMINISTRATION PROCEDURE AND INSTRUCTIONS

The following is a detailed description of the test administration procedure and instructions:

I. Meet the class.

- A. Be introduced as Mr. Swaner, from Utah State University, who is doing some very important research.
- B. Tell them that this research is concerned with children's feelings and that I will need their best cooperation and effort.
- C. See that everyone has a pencil.

II. Start with the Test Anxiety Questionnaire.

A. Read the instructions.

1. This takes care of the distribution of the answer sheets, the problem of confidentiality and explains the purpose of the test.

- a. "I'm going to be asking you some questions--questions different from the usual school questions, for these are about how you feel and so have no right or wrong answers. Let me give you an answer sheet." (Distribute the answer sheets.)

"Write your name at the top of the page where it says 'name.' Write your last name as well as your first. Also, by the word 'sex,' write a "B" if you are a boy and a "G" if you are a girl." (For fourth, fifth, and sixth grades) "Write the name of your school in the appropriate place, and where it says 'name of school attended last,' write the name of the school you attended last year.

"As I said before, I am going to ask you some questions. No one but myself will see your answers to these questions--not your teacher, your principal, or your parents. These questions are different from other questions that you are asked in school. These questions are different because there are no right or wrong answers. You are to listen to each question and then put a circle around either 'Yes' or 'No.' These

questions are about how you think and feel and, therefore, they have no right or wrong answers. People think and feel differently. The person sitting next to you might put a circle around 'Yes' and you may put a circle around 'No.' For example, if I asked you this question: 'Do you like to play ball?' some of you would put a circle around 'Yes,' and some of you would put it around 'No.' Your answer depends on how you think and feel. These questions are about how you think and feel about school, and about a lot of other things. Remember, listen carefully to each question and answer it yes or no by deciding how you think and feel. If you don't understand a question, ask me about it.

"Now let's start by everybody putting their finger on No. 1. Here is the first question. No. 1. 'Do you worry when. . .'" (The test items are then read.)

b. At the end, ask if there are any questions.

III. Distribute the TAT booklets, with the following comment:

A. "Let me give you one of these booklets."

B. Give directions as to how to fill in the information:
 "Please write your name in the space at the top of the page. By the word 'sex,' write 'B' if you are a boy, and 'G' if you are a girl. Now the date (tell them), your age, your name, and the name of your school. Now will you turn the page and fill in the information asked for at the top of the next page (age and sex)."

IV. Introduce the TAT.

A. "For this next test we will have to keep the room a little dark so that we can use the projector (set up before test begins) to show you some pictures. The room will still be light enough for you to read and write. Now listen carefully while I read the directions to you." (Read the following.)

This test is a test of Imagination. Imagination has been found to be one form of intelligence, so we would like you to do the very best that you can. I have some pictures here that I am going to show you, and I want you to write a story about each of the pictures. You will have 7 minutes in which to write your story. Notice, there is a separate page in your booklet for each story. If you need more room, use the other side of the page. In your story, tell who the people are, what they are doing, how they are feeling, what they are saying, and what is happening. Remember, this is a test of your imagination.

Do not merely describe the pictures. Write the best story you can. You may write any kind of story you please. Do you understand? Let me show you the first picture. Remember to tell a complete story about it.

- B. Present each of the seven cards for seven minutes in the following order: 1 M3 3BM 8G 10G 8B Blank
TAT MPT TAT MPT MPT MPT MPT
- C. At the end of each six minutes of card exposure, suggest that they have one more minute in which to finish their story.
- D. Allow a 3-minute break between cards 4 and 5.
- E. At the end of the test, collect the protocols.
- V. Introduce the CMAS by passing out the forms and saying: "Here is another test I would like you to take."
- A. After distributing the test forms, say: "Please write your name in the place at the top of the page. By the word 'sex,' write 'B' if you are a boy, and 'G' if you are a girl. Now the date (tell them), your age, your grade, and the name of your school."
- B. "Now listen carefully while I read the instructions at the top of the test paper. This is another test about you and how you feel. Read each question carefully. Put a circle around the word 'Yes' if you think it is true about you. Put a circle around the word 'No' if you think it is not true about you. All right, go ahead."
- C. When you have finished, collect the booklets.
- VI. Thank the class for their cooperation and leave.