

THE UNIVERSITY OF CALGARY

STORY SEQUENCE ANALYSIS OF THE THEMATIC
APPERCEPTION TEST: A REPLICATION STUDY
ON A SAMPLE OF STUDENT NURSES

by

GEORGE C. HARTWELL

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

DEPARTMENT OF EDUCATIONAL PSYCHOLOGY

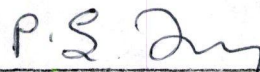
CALGARY, ALBERTA

OCTOBER, 1979

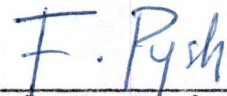
© GEORGE HARTWELL, 1979

THE UNIVERSITY OF CALGARY
FACULTY OF GRADUATE STUDIES

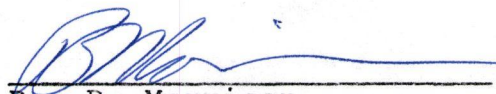
The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Story Sequence Analysis of the Thematic Apperception Test: A Replication Study on a Sample of Student Nurses", submitted by George Hartwell in partial fulfillment of the requirements for the degree of Master of Science.




Supervisor, Dr. P. S. Fry
Department of Educational
Psychology



Prof. F. Pysh
Department of Educational
Psychology



Dr. B. Morrison
Director of Psychology,
Calgary General Hospital



Dr. M. R. Rodway
Associate Dean, Faculty of
Social Welfare

October 18, 1979.

ABSTRACT

The purpose of this study was to replicate work on Arnold's Story Sequence Analysis method of scoring the Thematic Apperception Test using the hypothesis that there would be a positive relationship between SSA scores and academic achievement. In seven out of eight previous studies using Story Sequence Analysis in relation to school and college grades, there were successful scholastic predictions and discriminations.

A unique feature of this study, requiring scorers to learn this method from written materials, provided an evaluation of the objectivity of this scoring system and the viability of a psychologist learning SSA scoring from the manual. This procedure also guaranteed that the research in this study was done independent of Dr. Arnold, while in the majority of previous studies Dr. Arnold was involved in either the training of scorers or the scoring process.

The sample used in this study was from a secular Canadian institution, while all of the previous samples were American and many were from Catholic institutions. Subjects were 43 students in the first (N = 22) and second (N = 21) year of a two year psychiatric nursing training program. In contrast with most of the previous samples, these students were not

rigorously selected on the basis of intelligence or academic achievement. Each class was administered 13 TAT cards on a group basis. Scoring followed Arnold (1962) using +2, +1, 0, -1, -2 scores for each story.

Interscorer reliability coefficients between the two scorers were .50 and .71 for the first and second year class, respectively. There were some significant correlations ($p < .05$) between SSA scores and grades in individual subjects but no significant correlations between SSA scores and the final overall averages of the student nurses. The lack of validity of the SSA scores was related to factors of: (1) unreliability of grades and TAT scores, (2) low variability in grades, (3) high variability in intellectual variables, and (4) a possible contamination of the TAT administration.

From this, and previous studies, it was concluded that inconsistently adequate reliability levels of Story Sequence Analysis requires correction. The development of standardized training procedures based on written materials was suggested.

ACKNOWLEDGEMENTS

I would like to express my appreciation to the members of the thesis committee, Dr. P. S. Fry, Prof. F. Pysh, Dr. B. Morrison, and Dr. M. R. Rodway. Special thanks go to Dr. P. S. Fry for invaluable assistance in the revision process. Dr. Banner deserves acknowledgement for supervision of the research phase of the thesis.

Thanks go to Bob Buchanan and Mike Waye, and especially to Lee Dyer, for their involvement in the scoring of the TAT's.

I would like to mention Dr. D. Crouse who was my mentor in clinical psychology at Alberta Hospital Ponoka.

I am grateful to the staff and students at Alberta Hospital Ponoka for their cooperation with this study.

Thanks to my family for their patience and to the Cline family of Calgary for their support and hospitality.

I also wish to praise the Lord for the divine measure of grace, joy and peace that have been mine over this last year.

TABLE OF CONTENTS

	Page
ABSTRACT	iii
ACKNOWLEDGEMENTS	v
LIST OF TABLES	ix
Chapter	
I. INTRODUCTION	1
Murray's Approach to the TAT	2
Assumptions Guiding Traditional TAT Inter- pretation	3
The Need for a New Approach	4
A Successful Psychometric Approach to the TAT	5
The Purpose of the Study	6
Conclusion	8
II. REVIEW OF THE LITERATURE	9
Background to Story Sequence Analysis	9
The Role of the Story Outcome	10
Arnold's Theory of Achievement Motivation	11
Questionnaire versus Projective Measurement	13
The Initial Studies: Group Differences	14
The Use of Population-Specific Scoring Systems	15
The Use of the General Scoring System	17
Research since the Finalization of the SSA System	24

	Page
Research related to Academic Achievement .	25
The Experimental Studies with Story Sequence Analysis	32
Discussion of the Research Literature . .	36
III. METHOD	40
Sample	41
Administration of the Thematic Apperception Test	44
Scoring Procedure	47
The Scoring System	49
Training of Scorers	53
Data Analysis	56
Hypotheses	58
IV. RESULTS	59
Validity of the Motivation Index	59
Interscorer Reliability	63
The Characteristics of the Grades	64
Time Involved in Providing SSA Scores . .	69
Agreement Levels with Expert Scorers . . .	70
Summary of Major Findings	70
V. SUMMARY, DISCUSSION AND CONCLUSIONS	71
Rationale for this Study	71
Results	73
Discussion	75
Conclusions	78
Limitations of this Study	79

	Page
Implications for Further Research	80
REFERENCES	82
APPENDIX	85

LIST OF TABLES

Table	Page
1. ACE Intelligence Measures from Garvin (1960) .	20
2. Correlation between TAT scores, IQ Measures and Grade Point Average	21
3. Interscorer Reliability of Story Sequence Analysis in Studies since 1962	26
4. Validity and Research Results of Story Sequence Analysis since 1962	27
5. Multiple Correlations of IQ, SSA scores with GPA, and other Correlations involving IQ from Dulin (1968)	30
6. TAT Cards Administered	45
7. Correlations of Academic Achievement Measures with Arnold's Motivation Index, First Year Class	60
8. Correlations of Academic Achievement Measures with Arnold's Motivation Index, Second Year Class	61
9. Motivation Scores from Story Sequence Analysis of the TAT: Descriptive and Correlational Statistics	65
10. Means and Standard Deviation for Achievement Measures of the First Year Class	66
11. Means and Standard Deviation for Achievement Measures of the Second Year Class	67

Chapter I

Introduction

There is a need in clinical psychology for a relevant standardized method of scoring the Thematic Apperception Test (TAT). What is needed is an objective scoring system, that meets standards as a psychometric test and that, at the same time, produces information relevant to clinical assessment. Routine clinical use of the TAT involves scoring with norms subjective to the examiner (Klopfer & Taulbee, 1976). This practice cannot be supported by the research, and instead, reviewers (Anastasi, 1968; Vernon, 1964) have discouraged the use of projectives in any practical assessment situations because of lack of standardization of scoring, among other things. On the other hand, it has seemed to clinicians that available TAT scoring systems either lacked established reliability or validity or, if they do meet these standards, fail to yield enough relevant clinical data to justify the investment of time in scoring (Harrison, 1965). There is one method that appears to come close to fulfilling the desired objectives and that is Magda Arnold's (1949, 1962) method of Story Sequence Analysis.

Story Sequence Analysis is relevant to the clinician because the foundation of the method is a holistic approach to the story that looks at what the story says. A quantitative

measure, called the Motivation Index, is derived from this method and meets standards of reliability and validity.

The first step of analysis, using Arnold's method, is to abstract the meaning of the story--the story moral. This story statement, called an import, is the basis of clinical appraisal and of the global estimate of motivation level. Clinical analysis is based on the assumption that the underlying attitudes of the writer are expressed through the TAT story.

Objective scoring of story imports started when these imports were assigned positive or negative values, according to whether the import was more frequent in records of high achievers or low achievers. It is through the objective scoring of imports that the Motivation Index is derived. The objective scoring of imports and the Motivation Index is the basis for research using Story Sequence Analysis and for claims regarding its validity and reliability.

Murray's Approach to the TAT

The use of the TAT as a psychometric instrument may not reflect the intentions of its originator. The TAT was introduced to the field of psychology by H. A. Murray of Harvard. He especially recommended the technique as a preface to psychotherapy or psychoanalysis. His approach reflected both an interest in psychoanalytic theory, and a reaction against the increasing quantification being employed in assessment in the 1930's. The TAT, Murray believed, revealed inner

needs and drives which were not directly related to outward behavior. In the TAT manual (1943) he wrote:

The Thematic Apperception Test, familiarly known as the TAT, is a method of revealing to the trained interpreter some of the dominant drives, emotions, sentiments, complexes and conflicts of a personality. Special value resides in its power to expose the underlying inhibited tendencies which the subject, or patient, is not willing to admit, or cannot admit because he is unconscious of them.

Assumptions Guiding Traditional TAT Interpretation

The traditional assumptions guiding TAT interpretation have now been thoroughly questioned. Holt's (1961) paper criticized the proposition that TAT stories were formed as fantasy attempts at wish fulfillment much like daydreams. Holt's conclusions, in this matter, are now generally accepted (Arnold, 1962; Harrison, 1965).

There has also been general rejection of the more fundamental assumption that basic drives and unconscious needs find direct expression in TAT stories (Anastasi, 1968; Lazarus, 1961; Murstein, 1965; Vernon, 1964). This assumption had led to scoring approaches in which themes (affiliation, achievement, aggressiveness, etc.) were counted as a measure of the strength of the need or drive in the writer. Research suggests, instead, that the handling of the themes, by the writer, is more important than the predominance of a theme.

Research on Story Sequence Analysis supports the contention that how a theme is handled should be the focus of story analysis.

The Need for a New Approach

It may be that Murray did not intend that the TAT be used as a psychometric instrument yet it is now widely used in that capacity. In frequency of use, it is second only to the Rorschach among the projective tests (Brown & McGuire, 1976). Vernon (1964) thought it vital that we make up our minds whether to encourage or discourage the application of projectives in practical decision-making situations. His finding, and that of Anastasi (1968), based on an accumulation of data on standardization of administration and scoring procedures, adequacy of norms, reliability and validity, is that projective techniques cannot be recommended in a psychometric role for practical diagnostic and decision-making purposes.

Harrison's (1965) comprehensive review of research on the Thematic Apperception Test suggests that the TAT, combined with certain standardized scoring systems, can be used as a psychometric instrument. He reports positive results in both the practical and research uses of the TAT. Where negative results were forthcoming, Harrison suggested that "the likely cause of failure was some combination of inexperienced examiners, inappropriate scoring, unreliable or ambiguous criteria, and absence of preliminary research into

the characteristics of the criterion groups" (p. 584). He advocates focussed research in which the characteristics of the more and less proficient members of an occupation have been predetermined.

A Successful Psychometric Approach to the TAT

Positive results, from both the practical and research use of the TAT, suggest that, with certain standardized scoring systems, it may be possible to recommend the TAT for use as a psychometric instrument. A comprehensive review of research on the TAT, by Harrison (1965), found successful uses of the TAT in diagnosis, academic prediction, personnel research, military screening, assessing therapeutic outcome and research. Story Sequence Analysis, in particular, stands out, in Harrison's review, as a likely candidate for further research. Nine out of twenty-six successful practical applications of the TAT, mentioned by Harrison, involve Story Sequence Analysis.

In the area of prediction of academic achievement Harrison (1965) found that the TAT has been so uniformly successful that "one wonders why the work has not been replicated with larger populations with the eventual aim of practical application" (p. 582).

Story Sequence Analysis has produced useful data descriptive of a person's attitudes. Expectations of what actions lead to successful outcomes and what actions to failure have been revealed through the imports (story statements).

Information on the validity of the Motivation Index suggests a high and consistent level of prediction with regard to academic achievement (Annold, 1962; Brown, 1953; Dulin, 1968; Garvin, 1960; Honor, 1971; McCandlish, 1958; Snider, 1954) and personnel selection (Burkard, 1958; Petrauskas, 1958; Quinn, 1961; Steggert, 1961). The existing evidence (Ramirez, 1970; Smith, 1971) also suggests that the Motivation Index can be used with validity as a measure of therapeutic outcome.

The Purpose of the Study

The present study was designed to serve as a replication, in a Canadian setting, of research using the Motivation Index as an estimate of the motivation level of students. In this investigation, of Arnold's method of Story Sequence Analysis of the TAT, certain new features were added while other things were held constant.

The administration and scoring procedures were held constant by following, as closely as possible, Arnold's book, Story Sequence Analysis (1962). This book was used as a training and scoring manual.

Another constant in this study was the use of academic performance, as measured by year end grades, as the criterion of achievement. Eight of 15 previous studies, of Arnold's method of TAT scoring, used school or college grades as the achievement criterion.

The students who were subjects of this study were enrolled in a two year training course in psychiatric nursing.

This was the first sample of Canadian students and the first sample of student nurses to be studied using Story Sequence Analysis. This sample of students was from a secular setting whereas many of the previous studies were undertaken in a Catholic and separate school setting. This sample of students has been less vigorously selected for achievement and scholastic aptitude than the samples in at least five of the previous eight studies involving academic achievement.

To date, SSA scorers have learned this system as apprentice to someone who knew the system definitively. However, since it is not always practicable to learn a scoring system in the presence of an expert tutor, it was desirable to test this method with scorers learning only from the scoring manual and from consultation with one another. When Arnold published the SSA scoring system (1962), she suggested that it was easy to learn. In addition, Brown (1967, p. 25) comments that "The scoring system has been sufficiently finalized in manual form so that any psychologist can learn this projective methodology." By learning SSA scoring procedures from the manual it was possible to provide information as to the objectivity of SSA scoring procedures, and, as well, to provide information as to the viability of independent clinicians learning this method from the book.

Conclusion

In conclusion, reviewers have suggested that clinical psychologists abandon the use of projectives in a nonstandardized manner when the purpose of their use is to provide diagnostic or practical decision-making information. The interpretation of the TAT using subjective norms falls into this category and should be restricted to the limited use of an interviewing aid.

According to reports of the clinicians using this method (Arnold, 1949; Arnold, 1962; Brown, 1979) Story Sequence Analysis has produced useful clinical data about attitudes along with the Motivation Index. It is a holistic approach to the TAT story; is based on a cognitive theory rather than the fantasy drive theory and was developed through research focussed on the characteristics of target populations.

Information on the validity of the Motivation Index suggests a high and consistent level of prediction with regard to academic achievement and personnel selection. Reliability has been demonstrated but not consistently at the levels expected of a psychometric test.

The purpose of this thesis is to replicate work on Story Sequence Analysis with a new population and scorers learning from the book using academic achievement as criterion and standardized administration and scoring from Arnold's (1962) Story Sequence Analysis.

Chapter II

Review of the Literature

Background to Story Sequence Analysis

Arnold started the development of her new method of TAT analysis in Canada after the Second World War. As Director of Research and Training in the Psychological Services of the Department of Veterans Affairs, she was responsible for the training of psychologists to work in the psychiatric sections of veterans hospitals. Since there was no well-developed scoring system for the TAT, as there was for the Rorschach, she decided (Arnold, 1949) "to develop a method which is easy to teach and facilitates discrimination without making it superfluous by premature quantification."

In 1954, a series of research studies on Story Sequence Analysis began. At that time Dr. Arnold was professor of Psychology and Director of the Behavior Laboratory at Loyola University in Chicago. Together with graduate students she undertook research on normal people on the basis of blind analysis. This research laid the foundation for the scoring system published in Story Sequence Analysis (Arnold, 1962).

A Holistic TAT Scoring Method

A major innovation in Arnold's TAT scoring method was

the treatment of each story as a unit. Rather than scoring parts of the story separately, story elements were considered in relation to one another--as a gestalt. Arnold's contention was that "A story has a meaning which cannot be discovered from the individual themes [topics] into which it can be analyzed" (Arnold, 1962, p. 13).

Murray intended, through projective testing devices, such as the TAT, to foster holistic personality interpretation. Murray's (1943) need-press scoring system, however, required the scorer to put most of his time into the consideration of disconnected story elements. Most, if not all, of the scoring systems developed since 1943 have followed his lead in this.

The problem with the need-press scoring system was recognized by Murray (1943, p. 13), "To take up the hero and environment separately, as we have just outlined, involves the dislocation of the two fundamental elements of each concrete event." He suggested that the interpreter "put reality together again" by listing the most prevalent need-press combinations. He failed, at this point, to include an element that has become predominant - the story outcome.

The Role of the Story Outcome

The story outcome plays an indispensable role in Arnold's method. In Story Sequence Analysis the essential meaning of the story is expressed in that statement called the import. But the meaning of the story cannot be expressed

without considering the story outcome and the light it throws on the story action. Does the hero succeed or fail in his efforts? After an offense or crime does the hero get punished or does the hero "get away with it" without punishment or fateful consequence? These kind of questions become the essential focus of the interpreter in Story Sequence Analysis and not just a secondary consideration as in other scoring methods.

In Murray's Thematic Apperception Test Manual (1943, p. 12) the interpreter is asked to assess how much effort (force, determination) is manifested in the hero. The SSA scorer cannot consider this question because it separates the story character's personal effort from the story outcome. High achievers in the normal population (students, teachers, navy enlisted men) combine personal effort with a positive outcome and lack of effort with a negative outcome. Low achievers reverse this relationship: personal effort leads to failure and lack of effort brings success. Personal effort, or any other variable, cannot be considered alone in Story Sequence Analysis, but must be considered in relation to the story outcome.

Arnold's Theory of Achievement Motivation

Arnold is interested in overt behavior prediction. The scoring system for Story Sequence Analysis has been developed on the basis of research relating achievement behavior to TAT scores. More and less successful students,

more and less proficient teachers, navy men, executives and religious men have been studied. Prediction of achievement behavior based on SSA TAT scores would appear to be based on the following propositions (some implicit in Arnold's writings others explicit; some supported by research, others with no such support):

1) That the individual's personality includes a stable set of motivating attitudes. These motivating attitudes are habitual patterns of evaluation that are the basis for the person's action choices.

2) Motivating attitudes are expressed in the import derived from Story Sequence Analysis of the TAT or similar storytelling projectives.

3) That motivating attitudes are at a level of the personality that is assessed more successfully with the TAT than with questionnaire measures.

4) That motivating attitudes can be classified according to those most frequent in the stories of high achievers and those most frequent among low achievers.

5) That imports associated with high achievement can be given a positive score and those associated with low achievement can be given a negative score and the sum of these scores used to represent an individual's motivation level.

6) That there is a large range of motivation levels in a group of normal persons.

7) That motivation level and intelligence level are

potent variables in the prediction of an individual's achievement or performance level.

8) That life experiences, conscious decisions and therapeutic process are involved in the formation and alteration of motivating attitudes.

9) That thorough reading of the published scoring system and scoring procedures (Arnold, 1962), combined with group practice, are sufficient conditions for learning valid and reliable SSA scoring.

10) That SSA scores will be found valid for discovering positive and negative motivation in normal people in vocational or academic pursuits.

Questionnaire versus Projective Measurement. The proposition (#3) that motivating attitudes are measured more effectively with the TAT than with questionnaire personality measures was discussed by Arnold (1962, p. 38). She reports that both high and low achievers select positively scored outcomes when they are presented as multiple choice options. Honor (1970) found that SSA derived scores from the TAT were superior to scores from the California Study Methods Survey in distinguishing high achieving high school students from low achieving ones. Smith (1971) found SSA scores superior to selected scores from the Sixteen Personality Factor Questionnaire (16PF) and the Minnesota Multiphasic Personality Inventory (MMPI) as measures of therapeutic outcome.

In general, self-report and questionnaire measures of personality variables produce small and often contradictory correlations with achievement (Behrens and Vernon, 1978). The use of Story Sequence Analysis, on the other hand, results in "some strikingly successful scholastic and vocational predictions and discriminations" (Harrison, 1965, p. 579). Specific research comparing questionnaire and projective methods, as well as the more general trends, suggest that motivating attitudes are more successfully assessed using the TAT projective.

The Initial Studies: Group Differences

The initial studies found story imports that were present in one group and absent in another. Snider (1954) studied 40 high school seniors while Brown (1953) studied 40 sophomores. These students were in academic boys' separate schools where admittance was on the basis of entrance exams. Snider studied 20 boys from the top third of the class academically and 20 from the bottom third. Brown selected high- and low-achievers on the basis of predicted grades minus achieved grades. In both studies students were placed in matched pairs on the basis of equivalence of age, IQ and social background.

The studies of Brown and Snider discovered that significant differences can be found between the imports of high achievers and those of low achievers ($p = .05$). They also found consistencies in the imports characteristic of each

group. This was the initial evidence that there is a set of motivating attitudes generally characteristic of high achievers and a separate set characteristic of low achievers.

The Use of Population-Specific Scoring Systems

In the three studies to follow a large part of the sample was utilized in a preliminary study of matched pairs. On the basis of this large preliminary study a population-specific scoring system was established. Each of these objective scoring systems was then used in the blind analysis of the remaining protocols. The validity of the scoring system in discriminating high- and low-achieving persons was assessed in this second study. Interscorer agreement figures were derived as well.

Academic achievement discrimination. McCandlish (1958) was the first to demonstrate that an objective scoring system could be used to advantage along with Story Sequence Analysis. He used the material gathered by Brown (1953) and Snider (1954). Forty of these cases were used to develop a scoring system that was then tested on the remaining 40 cases. In 39 out of 40 cases it was confirmed that high achievers would have predominantly positive scores and low achievers predominantly negative ($p = .01$). Interjudge reliability was calculated between three judges using this scoring system with 97% agreement as to positive or negative scoring of stories. McCandlish found from his research

that his assignment of these protocols using his scoring system was more accurate than his assignment of them by clinical evaluation alone.

Discrimination of teacher effectiveness. The characteristics of effective and ineffective teachers were studied by Burkard (1958, 1962). Criterion groups were set up on the basis of pupil ratings on a teacher-rating scale. The original sample consisted of 300 teachers, all of whom were religious women teaching in parochial and private schools in the midwest. From this sample 50 high-rated teachers and 50 low-rated teachers were selected for study. Pairs of teachers were matched for age and intelligence and, as far as possible, grade, subject and type of school. Burkard developed an objective scoring system by scoring story abstracts--the imports--predominant in the protocols of high-rated teachers as "plus", and those of low-rated teachers as "minus".

In order to develop the objective scoring system, Burkard used 40 of the protocols for preliminary study. Then the completed scoring system was applied to the remaining 60 protocols. Out of these 60, 58 were correctly designated to the high- or low-rated group by one judge, 60 out of 60 were correctly designated by two other judges ($p = .001$), representing 100% separation of groups.

Interscorer agreement between pairs of scores, based on the positive/negative dichotomy, was: 97%, 97% and 94%.

Interscorer reliability is somewhat easier to attain when protocols at the extremes of motivating attitudes are scored and the middle-range eliminated. In Burkard's (1958) study the middle third of the protocols, the middle-range of teacher ratings, were not scored.

Navy men: offenders and non-offenders. A military application of Story Sequence Analysis was undertaken by Petrauskas (1958) who studied men enlisted in the navy. Thirty pairs of navy men were established--30 men in the brig awaiting court-martial and 30 men with no record of offences in the navy. Pairs were matched for age, IQ and length of service. Scoring criteria were established by studying 20 TAT records and the scoring standard applied to the remaining 40 records. The percentage of men, correctly assigned to the offenders or non-offenders, was 100% for one judge, 83% and 86% for the other two judges ($p = .01$). Agreement between pairs of scorers was assessed as: 82%, 80% and 80%.

The Use of the General Scoring System

The three researchers above, Burkard, McCandlish and Petrauskas, used scoring systems developed from the sample under study, and tested their validity with the same sample. In each case the middle-range of achievement was excluded, and the validity task was to assign a protocol to the high achievement or to the low achievement group.

A new phase of development began with studies that used the whole range of achievement. Prediction of ranking on the validity criterion was used instead of group discrimination. A general scoring system began to be used rather than a population-specific scoring system. Introduction of whole range studies, prediction studies and a general scoring system made the test of Story Sequence Analysis more rigorous. This new phase marked the transition of Story Sequence Analysis from a methodology for the study of particular populations to a general method applicable to a variety of populations.

These studies make use of a complex scoring system that provides two gradients of positive scores and two gradients of negative scores. This system was first worked out by Quinn (1959). In an initial form imports were scored 1, 2, 3 and 4, which corresponds to -2, -1, +1 and +2, in that order, in the published scoring system (Arnold, 1962).

Academic achievement prediction. Two studies investigated the general scoring system using school grades to provide criterion scores.

Garvin (1960) tested 100 college seniors (50 from a private women's college) using slide presentation of the TAT. Nine records were discarded because the test was incomplete or instructions not followed. Twelve records were selected using the students' Grade Point Average (GPA) to select students from the highest third ($N = 6$) and from

the lowest third ($N = 6$) of the sample. This preliminary study was used by Garvin to verify the general scoring criteria provided from the previous studies and to add new criteria that were needed for this sample.

Garvin (1960) found a significant positive relationship between GPA (Junior Year) and TAT scores, both for men ($N = 40$, $\underline{r} = .85$, $\underline{p} < .001$) and for women ($N = 39$, $\underline{r} = .83$, $\underline{p} < .001$). Scoring was done by the investigator using the early complex system (1, 2, 3, 4). Dr. Arnold acted as scoring consultant. No figures on interscorer reliability were reported.

In a second study of academic achievement prediction, Arnold (1962) studied seventh grade students using their final grades as criterion scores. Scoring criteria were added to the final scoring system (Arnold, 1962) from this study. There is no mention of a preliminary study and no report of interscorer reliability. There was a significant positive correlation between final grades and TAT motivation scores ($N = 51$, $\underline{r} = .75$, $\underline{p} < .001$).

Garvin used the American Council on Education Intelligence Test (ACE) in his study, and Arnold used the Otis IQ test. Statistics, shown in Table 1, indicate that the mean IQ of Garvin's sample was 124. In both studies moderate correlations between IQ's and TAT scores were found (Table 2). In multiple correlations of IQ and TAT scores, IQ contributes more in Arnold's study than in Garvin's.

Table 1
ACE Intelligence Measures from Garvin (1960)

	males (N = 46)	female (N = 45)
mean	125	122
standard deviation	18	20

Table 2
 Correlations between TAT Scores, IQ Scores and
 Grade Point Averages in Garvin's (1960)
 Study and Arnold's (1962) Study

	Garvin (males)	Garvin (females)	Arnold
<u>Correlation (r)</u>			
TAT with GPA	.85	.83	.75
TAT with IQ	.58	.47	.47
IQ with GPA	.63	.50	no report
<u>Multiple Correlation (R)</u>			
TAT, IQ with GPA	.87	.84	.84

It is likely that Garvin's students have been highly selected for intelligence, at college entrance, because the multiple correlation, involving IQ and TAT in the prediction of GPA, is only slightly higher than the correlation of TAT scores with GPA. Having been selected on the basis of college entrance examinations more of the variance in college grades is derived from motivation scores than from intelligence scores. According to these multiple correlations, TAT scores and I.Q. scores account for at least 71% of the variance in academic achievement of these samples. These multiple correlations match the best multiple correlations produced from batteries of tests to predict school achievement. The use of TAT motivation scores more than doubled the variance that could be accounted for, in this sample, by intelligence scores alone.

Vocational discrimination and prediction. In a study of federal executives, Stegert (1961) chose ten men that volunteered for executive development to compare with ten that did not. In general the men who had volunteered were rated "better than average" on job performance. The two groups were matched for civil service grade, years of service, type of position and education. Stegert used the early complex scoring (1, 2, 3, 4) and average score as an index of motivation. The results were as follows:

participants (N = 10) TAT scores	3.3 to 3.8,
non-participants (N = 10) TAT scores	1.7 to 3.2.

All but one of the executives expressed positive attitudes on the TAT yet there was no overlap of the scores of the two groups. This suggests that even where the range of motivation scores is reduced there can be successful vocational discrimination.

In a study of vocational prediction, Quinn (1961) studied 45 men who had taken temporary vows in a religious order and who were ranked by both peers and two superiors on "promise for religious life". The Motivation Index (MI), based on story sequence analysis of the TAT, produced correlations of $r = .59$ with peer-rankings, and of $r = .61$ with rankings by superiors. The rankings for promise for religious life were not, themselves, highly reliable, with a correlation coefficient of $r = .65$ between rankings by peers and rankings by superiors. So it is, in spite of an unreliable criterion that SSA scores were successful in vocational prediction in this case.

Discussion of the general scoring system. The studies with the general scoring system have demonstrated its validity for academic achievement prediction and for vocational prediction and discrimination. In three of these studies the middle-range of achievement has been included making the research relevant to practical decision-making. Since preliminary studies have been used, however, the complete independence of the research subjects from the sample on which the scoring criteria were developed could not be guaranteed.

Most of these studies were dissertations done under Dr. Arnold's supervision and, in some cases, with her assistance with scoring. Moreover, interscorer reliability of independent raters has not always been tested adequately.

Although the final four studies provide evidence for the validity of a general Story Sequence Analysis scoring system, none provide figures on interjudge reliability. For figures on the reliability of SSA scores, we are dependent on the studies by Burkard (1958), McCandlish (1958) and Petrauskas (1958). Unfortunately, these three investigators were using a simpler plus/minus scoring system; secondly, all of them made use of a preliminary study of their population and so produced a population-specific scoring; thirdly, all three used a split-group research design. The reliability of scorers using the extremes of the range of achievement and not the middle-range is misleading because, as Arnold notes (1962, p. 103), "The imports that are difficult to score are derived from stories of people whose motivation is neither consistently positive nor strongly negative."

Research since the Finalization of the Story Sequence Analysis Scoring System

Story Sequence Analysis was published by Arnold in 1962. The scoring system in that publication (-2, -1, +1, +2) may be considered the official final system and Arnold's book, the manual for scoring.

Interscorer Reliability. The interscorer reliability figures are displayed in Table 3. The difficulty of achieving high levels of scorer agreement is suggested by these figures. Less than half of the figures are over .80 which has been accepted as adequate for research purposes with n Achievement scoring of the TAT (Feld and Smith, 1958). Three of 14 are in the .70's which may be considered borderline acceptable but another five are below the .70 mark which puts them at unacceptable levels.

Validity. An overview of the research on the published scoring system is displayed in Table 4. Three of these studies use high school and college grades as the validity criterion. The other three studies are basically experimental studies.

Research Related to Academic Achievement

In the research on the published scoring system two studies (Brown, 1967, Dulin, 1968) were primarily involved in academic achievement prediction, one (Honour, 1971) involved primarily academic achievement discrimination, and one (Youngg, 1972) was involved in academic achievement prediction in a minor way.

Brown (1967). Brown's study involved administering the TAT and other measures to the complete freshman class of a small liberal arts college (113 men, 58 women). Brown was trained by Dr. Arnold over the course of a year.

Table 3
Interscorer Reliability of Story Sequence
Analysis in Studies since 1962

Study	Date	Interscorer Reliability	Comment
Fields	1964	81%	Reported in Brown (1967).
Brown	1967	.53 .64, .74 .80	Brown vs. Fields (external scorer). Inexperienced scorers vs. Arnold. Inexperienced scorers.
Dulin	1968	no report	Consensus scoring used.
Najjar	1967	.74	Reported in Ramirez (1970).
Honor	1971	.88, .92 ⁽¹⁾	Middle range excluded.
Ramirez	1970	.66, .70, .84	Experienced scorers with proximate training.
Smith	1971	no report	Consensus scoring used.
Youngg	1972	.58, .65 ⁽¹⁾ .89	First attempt trained scorers. After consensus rescoring 20 protocols.

(1) These studies simplified Arnold's published scoring system using a plus/minus system based on Arnold (1962).

Table 4
Validity and Research Results of Story Sequence Analysis
in Studies since 1962

Study	Population	Criteria	Result
Brown (1967)	college freshman	GPA freshman year	not significant
Dulin (1968) ^{ab}	college freshman	GPA first year GPA second year GPA third year GPA fourth year	$\bar{r} = .84$ $\bar{r} = .65$ $\bar{r} = .62$ $\bar{r} = .61$
Honor (1971) ^{abcd}	10th, 11th grade	GPA year end	$\bar{r} = .88$
Ramirez (1970)	counselling clients	Change of M.I. in therapy	significant $\underline{p} = .05$
Smith (1971) ^a	college seniors	Change of M.I. in therapy	significant $\underline{p} = .01$
Youngg (1972) ^c	college freshman	Increase in M.I. from indirect test clues	significant $\underline{p} = .01$

^aThese studies involve male subjects only.

^bThese studies involve a preliminary study.

^cThese studies involve a simplified plus/minus scoring system.

^dThis study excluded a large mid-range of achievement.

Brown failed to find a significant relationship between SSA scores and final freshman grades ($r = -.03$). The reliability figures (see Table 3) suggest low levels of agreement between an external scorer, not involved in scorer training, and Brown, a new learner of this method. The level of agreement between inexperienced scorers and an expert scorer (Dr. Arnold) is also low (.64, .74). The highest level of agreement is between the two scorers who trained together which suggests a high local level of agreement (unrelated to other scorers) can be achieved by training together.

Brown's (1967) study compared the construct validity of Need for Achievement (n Ach) TAT scoring and SSA TAT scoring. Construct validity was based on personality measures and on peer-ratings of characteristics derived from the theory and scoring practice of these two scoring systems. It was concluded that "Arnold's theory offers a better approach to the study of achievement by reason of its greater comprehensiveness in accounting for qualitatively distinct motivational variables" (p. 146).

In Brown's study, all correlations between SSA scores and peer-ratings were in the predicted direction while this was not true of n Ach scores. In addition, some successful predictors of GPA were weighted in the reverse direction from that predicted by n Ach theory. The greatest amount of variance in relation to GPA, was a strictly motivational factor clearly containing bi-polar positive and negative

directions. This bi-polar factor was predicted by Arnold's theory, which uses positive and negative scores, but was antithetical to n Ach scoring which has primarily positive weights. Brown's empirical research on the construct validity of these two methodologies provided evidence that Arnold's theory and scoring practice offer the best lead for the development of an adequate measure of achievement motivation.

Dulin (1968). Dulin's study is the first to predict grades in the years following the TAT administration. The stability of these correlations over the four years of college (.84, .65, .62, .61) suggests not only the possibility of long range prediction, but also the stability of motivating attitudes within the individual.

Dulin weakens the generalizability of his results, however, by the following procedures:

- 1) using only male students,
- 2) using a preliminary study (N = 10),
- 3) discarding an unrecorded number of protocols as unscorable (100 were selected for study from 700).
- 4) using consensus scoring procedures in which two scorers agree on a final score, and
- 5) involving Dr. Arnold as consultant to the scorers.

Dulin's study fails to report interscorer reliability figures.

The Henmon-Nelson Test of Mental Ability was used by

Dulin to provide an intelligence (IQ) measure. Multiple R's between IQ, TAT and GPA are recorded in Table 5.

Table 5
Multiple Correlations of IQ, SSA Scores
with GPA, and other Correlations
involving IQ from Dulin (1968)

Year	Correlation			
	Multiple R	IQ with GPA	TAT with GPA	IQ with TAT
1	.88	.44	.84	.24
2	.73	.45	.65	.20
3	.67	.34	.62	.16
4	.73	.48	.61	.12

The intelligence scores add a little to correlations with GPA, but the SSA scores, in this study as in most previous studies, add more to the prediction of achievement than any other measure. Correlations of IQ with TAT are small.

Honor found that SSA scores were a more accurate tool in discriminating high achievers from low achievers than the California Study Methods Survey (CSMS). On the basis of grades alone, 50 high school boys, with grade point averages (GPA) of 3.5 or higher, were tested as high achievers, while 50, with GPA 1.0 or lower, were considered as low achievers. The correlations between GPA and the TAT scores was .88;

between GPA and CSMS scores, .51. Only 2% were misplaced by the chosen "cut score" of the SSA scores while 25% were misplaced by the CSMS. Honor modified the SSA scoring system using a simple positive/negative scoring of each import. Honor provides evidence that SSA scores can be more effective than a questionnaire measure in predicting scholastic achievement.

The generalizability of Honor's results is weakened by the following points:

- 1) use of all male subjects,
- 2) use of a pilot study (N = 10),
- 3) use of a simplified plus-minus scoring system, and
- 4) failure to use the middle range of achievement.

Since a large mid-range of achievement is omitted stories are easier to score (resulting in higher reliability figures) and correlations are exaggerated (resulting in higher validity figures). In fact Honor's correlation between grades and TAT scores cannot be compared with the other such correlations due to this exaggeration of the correlation by the use of extreme scores only.

Youngg (1972). Youngg's study was primarily an experimental one, however, it did provide some figures on academic prediction. From the control group, which received a neutral lecture before the TAT administration, a significant correlation was found for a small sample of female students (N = 6, $\underline{r} = .88$, $\underline{p} = .05$). For males (N = 16) in the control group

no significant correlation was found.

The Experimental Studies with Story Sequence Analysis

The experimental studies reported here provide information on consensus scoring, the training of scorers and on factors that change motivating attitudes.

Consensus scoring. The use of consensus scoring between two or more scorers have been used by several investigators to reduce the unreliability of SSA scores. Of the six studies since the publication of Story Sequence Analysis in 1962, three (Dulin, 1968; Ramirez, 1970, Smith, 1971) use consensus scores as their motivation measure.

In Smith's procedure the TAT stories were scored independently by two scorers (Smith and Dr. Arnold) then discussed together to resolve scoring differences and arrive at a consensus score. Youngg's procedure involved two scorers who proceeded as above but a third scorer was also used as arbitrator if these two scorers could not resolve all discrepancies in scoring.

Youngg (1972) provided evidence that consensus SSA scores are more reliable than individual SSA scores. A group of 17 students were administered the TAT on two occasions, one week apart. The test-retest coefficient calculated from these TAT's by an individual scorer was .58. A second scorer also scored the TAT stories and together the two scorers derived consensus scores. The test-retest

coefficient using these consensus scores was .73. The difference between the two measures results from the greater reliability of the consensus scores.

If one uses the formula for the effect of unreliability on validity (Kolstoe, 1973) and makes estimates of any of the three unknown values (true test-retest reliability, reliability of consensus scores or reliability of individual scorers), limiting to values below .90, then the reliability of consensus scores lie between .81 and .90. The differences between the reliability of consensus scores and the reliability of the individual score is between .17 and .18. The reliability of the individual scorer would be between .64 and .72.

Training of Scorers. Ramirez (1970) used trained and experienced SSA scorers to provide a test of the reliability of scoring. Proximate training on 10 TAT protocols was done in consultation with Dr. Arnold. With this training and background interscorer reliability figures between pairs of scores averaged .73 (.84, .70, .66). This provides some evidence that even highly trained scorers are not able to produce SSA scores with reliability levels which are consistently above the .80 standard.

Youngg (1972) provides information on the increase of reliability with training. Two scorers scored 20 protocols (a different set each) to compare with her scoring. Interscorer reliability average was .61 (.65, .58), which is inadequate

for a psychometric device because of the large amount of error introduced into the observed scores (Kolstoe, 1973). Further intensive training was provided to Youngg and one of the scorers by rescoring 20 sets of TAT stories in a conference together in order to resolve scoring differences and arrive at consensus scores. The effect of this training was observed when these two scorers independently scored a further 20 protocols and interscorer reliability coefficients increased from .65 to .89.

Youngg demonstrated that intense training between two scorers greatly increases agreement levels between these two scorers. However, Brown (1967), found that high reliability levels, between two scorers who trained together, were not matched by agreement levels between these scorers and other scorers. It may be that intensive consensus training between two scorers creates a high local level of agreement, not shared with other SSA scorers.

Stability and change of SSA scores. All three experimental studies dealt with stability and change in SSA scores. Youngg (1972) investigated the influence of both background knowledge and recent events on SSA scores. Background knowledge of the TAT did not effect scores. Instructions to "try to make a good impression" resulted in significantly lower group means ($p = .01$). Indirect test clues provided in a guest lecture by the administrator on the characteristics of high achievers, resulted in significantly higher ($p = .01$)

group means than a "neutral" guest lecture on sensitivity groups. Youngg's study provides strong evidence that recent cognitive input and administrative instructions can alter SSA scores.

Evidence that therapeutic intervention results in stable changes in SSA scores and motivating attitudes is provided by Smith (1971) and Ramirez (1970). Smith established a control group of 35 for 35 men undergoing brief group sessions of disclosure-confrontation therapy (Tyrell, 1971). The subjects were normal college students. No significant change in the mean SSA scores of the control group was found ($p = .05$). The experimental group were tested one month after the session and significant increases in TAT scores from pre-test to post-test ($p = .01$) were found.

Ramirez (1970) used adults as the experimental group, undergoing client-centered therapy and those awaiting therapy as the control group. In his study there was a significant decrease in the SSA scores of the control group ($N = 26$, $p = .05$). The experimental group demonstrated a significant increase in SSA score after therapy ($N = 30$, $p = .05$). Experimental subjects who were rated as successful by their therapist ($N = 15$) demonstrated significant increase in Motivation Index from pre-therapy levels. This increase was significant both at post-test ($p = .001$) and at a six-month follow-up ($p = .05$).

Youngg demonstrated that administrative events effect SSA

scores. The importance of the findings of Ramirez and Smith is to demonstrate (1) that motivating attitudes of an individual undergo systematic and stable changes, (2) that SSA scores of the TAT can be more effective than questionnaire measures in assessing changes occurring in counseling, (3) that Story Sequence Analysis of the TAT can be a sensitive measure of therapeutic change, and (4) that changes in a person that a therapist rates as positive (greater adjustment, mental health, success in therapy, etc.) are associated with changes of the Motivation Index in the positive direction.

Discussion of the Research Literature

Story Sequence Analysis scores in the form of the Motivation Index or an equivalent are, on the basis of some of the research received above, more highly predictive of achievement than practically any other previously published study based upon a single methodology. The results of the research demonstrate that Story Sequence Analysis is valid for both prediction and discrimination of achievement in 14 out of 15 studies reviewed. The promise of this objective method of scoring the TAT, for practical decision-making and clinical work, seems bright.

The transition from the research situation to the practical setting, however, is not complete. Certain aspects of the research limit the generalization of the results to the practical setting. The selection of subjects,

by Honor (1971), and the selection of protocols by Dulin (1968) removes their research conditions a step away from the real situation. The use of a preliminary study means that the complete independence of the research subjects from the sample on which the criteria were developed cannot be guaranteed. Even in research with the finalized scoring system, two researchers (Dulin, 1968; Honor, 1971) use a preliminary study.

The use of a simple plus/minus system by Honor (1971) and Youngg (1972) limits the applicability of their results to the final published system which uses +2, +1, -1, -2. Dr. Arnold's involvement in scoring (Dulin, 1968; Smith, 1971) or in training of scorers (Ramirez, 1970) set these situations apart from that of the clinician or clinicians learning this projective methodology from the manual. The consensus scoring procedure, used by three of the researchers (Dulin, 1968; Smith, 1971; Youngg, 1972) is not practical for clinical settings. These factors argue for research conditions more closely approximating the practical situation in which the clinical psychologist would learn and use this method.

The evidence on the reliability of scoring suggests that reliability levels fall below those reached by Need for Achievement scoring of the TAT (.80). The careful judgment required in formulating the import, and the skill and concentration required in scoring the import, because of the complexity and size (45 pages) of the scoring system, make it

easy to understand why lower agreement among scorers is being found. What to do about this problem is another question. Lengthy and intensive training in Youngg's (1972) research succeeded in raising agreement levels to the .88 level. Further research and development on the training of scorers is needed, however, because of the possibility that Youngg's procedure produces reliable but not valid SSA scores.

The limitations in the applicability of the research results in no way detracts from the validity of the information provided by the researchers. Smith (1971) and Ramirez (1970) have provided valuable new information on the validity of SSA scores vis-a-vis counsellor's ratings of success in therapy on theoretical propositions regarding the stability and change of motivating attitudes, and on the practical value of SSA scores in assessing therapeutic change. Honor (1971) and Smith (1971) have demonstrated the superiority of SSA scores to questionnaire measures for academic achievement discrimination and discrimination of therapeutic change respectively. Dulin (1968) has provided a valuable replication of Gavin's striking results in the prediction of college achievement. Dulin (1968) demonstrated long term prediction using SSA scores. Brown (1967) provided information on the theoretical adequacy of Arnold's theory and scoring practice in a test of its construct validity and in a comparison of this theory of motivation with that derived from Need for Achievement scoring. Brown concluded that

Arnold's theory offers a better approach to the study of achievement motivation. "Arnold's methodology offers the best lead for the further development of a single, more adequate empirical measure of achievement motivation" says Brown (p. 150). Certain methodological and technological problems must be solved but the possibility of a valid, relevant and reliable method not only for scoring the TAT, but also for assessing motivating attitudes is available in Story Sequence Analysis.

Chapter III

Method

The Story Sequence Analysis method of scoring the TAT has been found useful for academic and vocational predictions and discriminations with certain populations in the American setting. Although the initial development of this method was for clinical use in a Canadian setting, there have been no studies of achievement prediction using a Canadian sample. In addition, many of the previous studies have been undertaken in Catholic institutions. Where SSA scores have been used for the prediction of academic achievement, the samples have been from student populations highly selected, by their institution, for academic achievement and intelligence. By way of contrast, this study was done in a secular Canadian institution where the students were not rigorously selected on the basis of high school record or academic aptitude.

In this study, administration of the TAT, training of the scorers and scoring procedures followed closely those outlined in Arnold's book Story Sequence Analysis (1962). Unlike many previous studies, Dr. Arnold was not personally involved in supervising the research, training the scorers or in the scoring.

Sample

This study used students enrolled in a two-year program for the training of psychiatric nurses as a sample. This sample is different from other samples which used the Motivation Index for academic prediction. Previous studies have involved grade seven students (Arnold), high school students (McCandlish, 1958; Honor, 1971) or college students (Brown, 1967; Garvin, 1960; Dulin, 1968). The student nurses in this sample were not as well screened by their institution for academic standing and aptitude as may be the case in the college samples. In fact, students were admitted to this two year training program who did not have the requisite high school transcripts or who did not fulfill the requirements of five Alberta grade twelve subjects.

One advantage of this sample, for study with Story Sequence Analysis, is that the School of Nursing policy continues to allow for the use of psychometric testing in the evaluation of applicants. In the past, the School of Nursing has used psychometric testing to screen applicants but did not find the results satisfactory according to the report of the staff at the School. Thus, the assumption is that if the Motivation Index is found valid as a predictor of the achievement of these nurses then it could be put to practical use by the School for the screening of applicants.

Locale. The locale for this study was the Alberta Hospital School of Nursing, Ponoka, Alberta. Two such Schools

of Nursing are operating in Alberta. They work in cooperation with the two provincial psychiatric facilities. Thus students are involved in clinical experience on the psychiatric wards and serve as Nurse Interns during the summer. Students are allowed to earn an annual remuneration of \$5,000.

Courses and evaluation. The first year of the academic program included eight subjects while, in the second year, there were five. The eight first-year subjects of instruction were:

1. Anatomy and Physiology,
2. Nursing I,
3. Pharmacology,
4. Introductory Psychology,
5. Introductory Sociology,
6. Introductory Psychopathology,
7. Introductory Family, and
8. Psychiatric Nursing I.

The five second-year courses were:

1. Medical-Surgical Nursing,
2. Psychiatric Nursing II,
3. Family Dynamics,
4. Psychopathology, and
5. Theories of Personality and Systems of Psychotherapy.

Evaluation in these courses was based on assignments, term tests, and final examination. For each student a final average was calculated in each subject and from these subject

averages a final overall average was calculated. Provincial comprehensive examinations were written by students completing their second year and included the following examination areas:

1. Medical-Surgical Nursing,
2. Psychiatric Nursing, and
3. Social Sciences

Subjects. At the time of testing, there were 50 students enrolled in the psychiatric nursing course at Alberta Hospital, Ponoka. Five of these students, all from the first year class, were absent on the day of testing. Two students were dropped from the sample because less than 10 of the TAT stories they wrote were scorable. Of the 43 students with scorable TAT protocols, 22 were first-year students and 21 were second-year students. These 43 students were the subjects for this study.

The heterogeneity of the subjects can be seen on a variety of variables. Fourteen were male and 29 female. Ages ranged from 19 to 42 with mean ages of 21.9 years and 23.6 years, for the first and second year classes, respectively. Five were married and 38 single. Seven were local residents, 20 had applied from other parts of Alberta, while 16 were out-of-province applicants.

The educational background of the subjects was varied. There were seven students who were admitted on adult privileges (i.e., they did not fulfill the requirement for a

minimum of 4 grade 12 subjects). Of the remaining 36 subjects, 29 received a high school diploma, 5 graduated with senior matriculation, while 2 were admitted on the basis of university transcripts.

Administration of the Thematic Apperception Test

The testing was sanctioned by the director of the School of Nursing and scheduled for April, 1976. The day of the testing was during the last week of classes before final exams. The first-year class was tested in the morning and the second-year class in the afternoon of the same day. Each class was tested, as a group, in their regular classroom setting.

Total time for test administration per class was 2½ hours. This time included 15 minutes for introduction and set-up and approximately 75 minutes to 120 minutes for the writing of the stories, depending on the time needed by the individual student to finish the writing of the stories.

Test materials. Sets of TAT cards were distributed to the students. Enough sets were available so that one set could be shared among two students. Each set contained 13 TAT cards which were presented in a specific order.

The following is Murray's designation for the TAT cards administered in the order of administration: 1, 2, 3BM, 4, 6BM, 8BM, 9BM, 10, 11, 13MF, 14, 17BM, and 20. A description of these cards is available in Table 6. Thirteen cards were

Table 6
TAT Cards Administered

Card No.	Description
1	A young boy contemplating a violin on a table in front of him.
2	Rustic farm scene: man with horse drawn plow in the background, middle aged woman looking on and, in the foreground, a young woman holding two books
3BM	A huddled form of a young boy, or girl, knees on the floor and slouched over a couch resting his head on his arm. An ambiguous object, possibly a revolver, on the floor.
4	A woman clutching a man's shoulders who is turned from her, drawing away with an intense look in his face.
6BM	An elderly woman looks out a window turned away from a young man standing looking downward, with hat in his hands.
8BM	A young adolescent boy looking out of the picture. A rifle leans to one side of him and a somewhat obscure lantern-lit surgical operation can be seen behind him.
9BM	Four men in overalls asleep on the ground.
10	A woman resting her head on a man's shoulder.
11	An ambiguous drawing with dragon-like figure protruding from a rock-face and some unidentifiable figures crossing a bridge over a chasm, below.
13MF	In a bedroom a man stands with downcast head while prostrate on the bed behind him is a partially uncovered form of a woman.
14	A silhouette of a human figure looking out a lighted window.
17BM	A nude man climbing a rope.
20	A dimly illumined figure leaning against a lamp post in the dark of night.

administered so as to allow three stories to be omitted or be unscorable before omitting a protocol. (Arnold (1962) considers 10 the minimum number of stories necessary for the calculation of the Motivation Index.) The TAT cards selected for administration were chosen from among those recommended by Arnold (1962) and those used by previous researchers of Story Sequence Analysis (Burkard, 1958; Garvin, 1960; Petrauskas, 1958; Steggert, 1961). Arnold uses the M pictures of Murray's TAT set for both men and women because these pictures are dramatic and lend themselves easily to imaginative exploration by both sexes.

Instructions. The instructions for administering the TAT were adopted from Arnold (1962, p. 49). Before the first story was written the first set of instructions was given. Seven minutes were allowed for the writing of the first story then the second set of instructions was given. Seven minutes were allowed per story so it was after 14 minutes that the final or third set of instruction was given. The instructions read as follows:

Set One

This is a test of your creative imagination. You will see thirteen pictures, one after another. As you look at each picture, write as dramatic a story as you can about it. Tell what has led up to the scene shown in the picture and what is happening now. What are the thoughts and feelings of the people in the picture? What will be the outcome?

Since we are interested in your creative imagination, be sure to write a story with a plot and an outcome. Do not just describe the picture. Try to write a story and not a piece of conversation.

You will have seven minutes for each story. Be sure to write something about each picture. If you can't think of anything for one of the pictures, write that down, too.

Set Two

Remember that we are interested in your creative imagination, and be sure to tell a story with a plot and outcome. Tell what has led up to the scene in the picture, what is happening now and how it will end. Write a straight story, not a piece of conversation.

Set Three

Remember, we want to know what goes on in the picture, what led up to it, what is happening now, and how it will end. Write a straight story, not a piece of conversation.

Scoring Procedure

Scoring TAT protocols using the Story Sequence Analysis procedure involves three steps. The procedure used in the present research involved only the first two steps: the first step was formulating the story imports; the second step was scoring the story imports. A third step, used in clinical evaluation, involves analyzing the story imports in sequence for a connected statement of the storyteller's principles of action and dominant concerns. The third step is not necessary for deriving motivation scores so was not used in this study.

Step one: formulating the story import. The first step of Story Sequence Analysis is to abstract the story import. The import is a short but complete expression of what the story is saying. Through the import the significant

meaning of the TAT story is stated in general form.

The following are important guidelines for the formulation of the story import:

1) The import must be objective in the sense that it is abstracted as accurately as possible from the actual story without adding any kind of interpretation.

2) The import is written from the point of view of the main character in the story.

3) The import is formulated in the second or third person but not in the first person.

4) The import abstracts or generalizes from the specifics in the story. If it is too specific it becomes simply a summary, failing to convey the meaning or significance of the study.

5) The import must convey the essential meaning and preserve the individual note of the story, the nuances of the story. The latter is done by using the actual words of the story, in some cases.

6) Each import is written so as to link the story imports together and so indicate the development in sequence of the thoughts expressed in the stories. It is especially important to bring out this sequence when the purpose of analysis is to obtain a clinical evaluation.

Step two: scoring the story import. Quantitative scoring of the import is done by using the scoring system. Arnold's scoring system (1962, pp. 226-271) catalogues

criterion imports. To use this system, the scorer matches the import to be scored with the import in the scoring system with the nearest meaning. The scoring process is aided by general principles that guide the assignment of scores and by the categories and headings that organize the scoring system. Once the criterion import corresponding to the story import is found, the import is scored as +2, +1, 0, -1 or -2 according to the designation of the criterion import.

If, in spite of every effort, the import does not seem to fit into any scoring category then the story should be consulted again. The scorer must ensure that important details of the story have not been neglected. If, however, the import includes everything in the story necessary for accurate scoring then the scoring system is used as a general guide to correct scoring. General principles have emerged from the empirically derived scoring system. These general principles can serve as a guide for assigning the appropriate score.

The Scoring System

Material in the scoring system (Arnold, 1962, pp. 226-271) is organized under four main themes represented by four categories in the scoring system. Under these categories 15 main headings further specify the theme. Under each of the 15 main headings the material is found under the designated scoring whether +2, +1, 0, -1 or -2. Subheadings under the designated scoring further classify criterion

imports according to how the theme is resolved.

The four scoring categories. There are four scoring categories: I. Achievement, success, happiness, active effort (or lack of it); II. Right and wrong; III. Human relationships; IV. Reaction to adversity. These include all the imports likely to be scored in stories.

Category I includes success in business, profession, work, finances and also happiness and success and failure in wider senses and from effort of every kind. This category includes attitudes to work, rest, recreation, success and failure. Goals and values are considered under this category including adaptability as to goals and means towards goals.

Category II, right and wrong, includes well-intentioned, reasonable, constructive or responsible action and its opposite: ill-intentioned, impulsive, harmful, destructive, or irresponsible action. Actions considered in this category include: wrongdoing, duty, fighting injustice and accidentally harmful action. Intentions are also scored including: temptation to do wrong, revenge, negative emotions and wrong attitude. Material on punishment and its effects is found in Category II.

Category III includes actions and attitudes related to human relationships. Under this category is material related to love, marriage and friendship, also, hate and bad relationships. The story that focusses on attempts to

help, encourage, pressure, advise or influence another person is scored under Category III.

Category IV, the adversity category, covers loss, harm, danger, terror, separation, disappointment and difficulties.

There are 15 main headings that provide a further breakdown of the story theme under each category. A description of these 15 headings is provided in the Appendix.

Within the 15 main headings, criterion imports are arranged under the different scores. These scores range from very positive to very negative: +2, +1, 0, -1 and -2. The zero (0) score is used almost exclusively for unscorable stories. An unscorable story means that the subject omitted that card or wrote a card description. The failure to write a story with a plot makes it very difficult to score using Story Sequence Analysis.

Positive two score (+2). The most positive imports scored +2, portray, in general, overt and positive action. (This has been discovered empirically from the records of high achieving students, well-adjusted Navy men, good teachers and ambitious executives.) For example, in the event of the death of a loved one the central character, by determined effort, overcomes the loss, perhaps even finding the strength to comfort others. A story in which wrongdoing results in just punishment would be found in the +2 scoring. Another example would be a story in which success follows from personal effort.

Positive one score (+1). The +1 import portrays activity that is positive but not so overt, or, is overt but not so positive. For example, a story might concern a student who takes care to get enough sleep and plans ahead and, therefore, is successful. Such a story emphasizes planning and good management. In this way it is positive but not as overtly so as a story that emphasizes the amount of work done by the student. In another context the central character overcomes adversity with some personal initiative but the story emphasizes worrying. Such a story is less positive because of the emphasis on worrying.

Negative one scoring (-1). Found in many of the -1 criterion imports is a passive attitude and lack of any positive action. External forces bring success; others help one overcome adversity. Positive goals are accomplished through such means as: fate, the passage of time, outside assistance, worrying, prayer, resolutions and dreams becoming reality.

If the right action is done for the wrong motives, this, too, is scored -1. Some examples: 1) work is done under pressure, 2) success is achieved despite negative attitudes, and 3) good relationships are maintained out of fear.

Negative two scoring (-2). The -2 import is actively negative, overtly malicious or impulsively destructive. If,

for example, the central character reacted to adversity with desperate action or total despair, the import would be scored -2. Among -2 imports good relationships are not durable and are broken without any effort at reconciliation. Work is evaded.

One also finds, in -2 criterion imports, a reversal of the relationship between action and outcome that is found in the +2 imports. For example, active effort leads to failure and wrongdoing succeeds if one is clever.

Training of Scorers

Although guidelines for abstracting and scoring the story import are provided in Arnold's book (1962), she strongly recommends that formulation of imports be a group project until considerable facility has been achieved. Other researchers (Youngg, 1972) have had scorers conference scoring discrepancies, as a form of training, to increase interscorer reliability. Both these procedures were involved in the training of scorers in this research.

The steps used in training were as follows:

- 1) Required reading in Story Sequence Analysis (Arnold, 1962) of chapter 4, the import and the sequence; chapter 5, positive and negative imports; chapter 7, using the scoring criteria; chapter 8, scoring difficulties and Appendix A, the scoring system.
- 2) Group training in abstracting and scoring the import.
- 3) Independent scoring of training protocols.

4) Conferencing scoring difference with a view to resolving differences and arriving at a consensus score.

5) Repeating steps 3 and 4 until interscorer reliability reached 70% plus-minus agreement on individual stories, or, if enough sets are scored, a .70 correlation between the Motivation Indices assigned by one scorer and those assigned by the other scorer.

Initial training program. A pilot project provided valuable information on the training process, on interscorer reliability levels and on the level of agreement between the scorers in this project and experts in Story Sequence Analysis (S.S.A.) scoring.

The initial training program involved three scorers. The first scorer, the investigator, was self-taught in the use of Story Sequence Analysis and had used it previously for psychiatric patient assessment. The second scorer, a graduate student in Psychology (Doctorate program), had no experience with TAT assessment. The third scorer, a graduate student in Education Psychology (doctorate program), had experience with the use of the TAT but had not encountered Story Sequence Analysis of the TAT.

First trial round of scoring. The three persons in the initial training programs spent a weekend (15 hours) in group training. After group training, independent scoring of 59 training stories was done by each scorer. Interscorer

agreement was calculated by a plus-minus agreement on individual stories by pairs of scorers. Agreement levels were 57%, 60% and 65% for scorer pairs 1:2, 1:3 and 2:3, respectively.

Second trial round of scoring. Two of these scorers, the first and the third, were involved in a second round of scoring. After scoring discrepancies were resolved they scored a further 51 training stories. Plus-minus agreement between this pair of scorers increased from 60% in the first round to 72% in the second round.

Consensus scores were arrived at between these two scorers. These consensus scores were the result of discussions aimed at resolving differences in scoring. Consensus scores on these 51 stories served as a scoring standard for the training of the fourth scorer.

Final training program. The fourth scorer paired with the first scorer formed the team for scoring the protocols of the 43 subjects in the sample. The fourth scorer had undergraduate training in the social sciences plus training in community leadership. She had no previous experience in scoring the TAT. The first phase of this scorer's training consisted of reading Story Sequence Analysis, instruction and supervised practice in abstracting and scoring imports. In the second phase of training, over the next eight weeks, approximately 20 stories per week were scored by this scorer

and conferenced with the first scorer.

During this training program the fourth scorer scored 120 stories. There was 70% overall plus-minus agreement between scorer four and the criterion (either the consensus scores or scorer one). A final check on the level of inter-scorer reliability was made, before all the subject protocols were scored. A sample of 10 sets of stories from the first class was selected at random and scored by both scorers (one and four). A Spearman rank order correlation of .77 was obtained which suggested that interrater reliability was sufficiently high to score the remaining protocols.

Data Analysis

As a result of the scoring procedure described above each TAT story was assigned a raw score from +2 through to -2. The next step was to add together the raw scores for each protocol. The sums of the raw scores provided some indication of motivation level. Positive sums indicated positive motivation, negative sums indicated negative motivation and sums near zero suggested something in the middle. The difficulty with using the sums of the raw scores as the index of motivation level is that sums from protocols of different length are not comparable. For example a +20 from a set of 10 stories would mean something different from a +20 from a set of 13 stories.

In order to make sums of raw scores comparable these sums were transformed into the Motivation Index. The trans-

formed scores may be represented by the formula $\frac{(S + 2N)}{2N} \times 100$ where S is the sum of raw scores and N the number of scorable stories. Since S varies between $-2N$ and $+2N$ the Motivation Index varies from zero to 200. A raw score sum of zero would be transformed to a Motivation Index of 100. A score of +20 from 10 stories would be $\frac{(20 + 20)}{20} \times 100 = 200$ while a +20 from 13 stories would be $\frac{(20 + 26)}{26} \times 100 = 177$.

It was assumed that the distributions of student grades and motivation scores were near enough to a normal distribution and that both student grades and motivation scores were derived from a scale near enough to an interval scale to use parametric statistics. A decision was made to use the product moment correlation coefficient (r). (Actually the raw scores of +2, +1, 0, -1 and -2 cannot be shown to be separated by exactly equal intervals and thus it would be more precise to indicate that an ordered metric scale is involved. However, Abelson and Tukey (1959) argue that data from an ordered metric scale can be treated as though it were interval data, and Labovitz (1970) argues that the small error involved in the use of interval statistics in such a case is more than offset by the gain from using more clearly interpretable statistics.)

The significance level for testing hypothesis was set at the .05 level. Three motivation scores were calculated for each subject. One was derived from each of the two scorers. The third motivation score was derived from the

first two and was the arithmetic mean of the motivation scores assigned by these two scorers.

The main measure of academic achievement will be the final average assigned each student by the School of Nursing. Secondary measures of academic achievement will include the final average assigned each student in the individual subjects of instruction and the marks assigned through the provincial examinations of the second year students.

Hypotheses

The main hypothesis is that there will be a significant positive relationship between the SSA scores, from the TAT protocols of the student nurses, and the final overall averages assigned these students by the School of Nursing.

Secondary hypotheses are the following:

1) that there will be a significant positive relationship between SSA scores derived by new learners of this method and final marks in individual subjects,

2) that there will be a significant positive relationship between the SSA scores derived by new learners of this method and grades on the provincial comprehensive examinations, and

3) that there will be a significant positive relationship between the average motivation score (arithmetic mean of the SSA scores of the two scorers) and the main and secondary measures of academic achievement of the psychiatric nursing students.

Chapter IV

Results

The purpose of the present study was to investigate Arnold's (1962) Story Sequence Analysis method of scoring the TAT. The relationship between Arnold's Motivation Index and measures of academic achievement was of major interest. Interscorer reliability was another point of interest. As an aid in estimating the feasibility of using SSA scoring for clinical assessment or practical decision-making, data and observations from the scoring and training of scorers were provided.

In the analysis of the results, the SPSS computer package was used to make the calculations. Pearson product moment correlations were used for correlations throughout this chapter.

Validity of the Motivation Index

The validity of the Motivation Index was the central point of interest. The correlations between the Motivation Index (derived from SSA scoring of the TAT) and academic achievement measures are displayed in Tables 7 and 8.

Validity with respect to final overall averages. The primary criterion of academic achievement was the student nurses' final overall averages. The observed correlations,

Table 7
Correlations of Academic Achievement Measures
with Arnold's Motivation Index,
First Year Class

Achievement Measure	Source of Motivation Index		
	first scorer	second scorer	average score
Full Class (N = 22)			
Final Overall Average	.21	.21	.24
Anatomy and Physiology	.14	.14	.16
Nursing I	.16	.30	.27
Pharmacology	.26	.37*	.36*
Psychiatric Nursing	.20	.00	.11
Intro. Psychology	.27	.36*	.36*
Intro. Sociology	.18	.19	.22
Intro. Psychopathology	.11	.05	.09
Intro. Family	.18	-.01	.10
Follow-up Group (N = 17)			
Second Year			
Final Overall Average	.20	.21	.24
Provincial Examinations:			
Medical-Surgical Nursing	.17	.21	.22
Psychiatric Nursing	-.06	.10	.02
Social Sciences	.37	-.01	.21

Note: the follow-up group consists of the 17 students from this class who completed their two year program.

* $p < .05$

Table 8
 Correlations of Academic Achievement Measures
 with Arnold's Motivation Index,
 Second Year Class

Achievement Measure	Source of Motivation Index		
	first scorer	second scorer	average score
Final Overall Average	.10	.20	.16
Medical-Surgical Nursing	-.09	.17	.04
Psychiatric Nursing II	.36*	.44*	.43*
Family Dynamics	.05	.20	.13
Psychopathology	.00	-.06	-.03
Psychotherapy	.13	.00	.07
Provincial Examinations:			
Medical-Surgical Nursing	.19	.37*	.30
Psychiatric Nursing	-.24	-.33	-.31
Social Sciences	.05	-.10	-.03

Note: N = 21

*p < .05

between the Motivation Index and final overall averages, were not significant at the $p = .05$ level for either class. For the first year class the observed correlations between SSA scores and final overall averages were .21, for both scorers, and .24, for the average of the SSA scores assigned by the two scorers. The null hypothesis was accepted for the first year class.

For the second year class the observed correlations between SSA scores and final overall averages were .10, .20, and .16, for the first scorer, the second scorer, and the average score, respectively. Since none of these correlations were significant at the $p = .05$ level, the null hypothesis was accepted for the second year class as well. In none of the relationships displayed in Table 7 or Table 8 does the average score display a significant advantage.

Validity: secondary measures. It was found that there were significant correlations between motivation scores and the secondary measures of academic achievement. For the first year class there were four correlations significant at the $p = .05$ level (see Table 7). Pharmacology and Introductory Psychology final averages were positively correlated with motivation measures (Motivation Index) derived from the second scorer and the average of the two scorers.

The follow-up group, referred to in Table 7, was comprised of the 17 subjects from the first year class that completed the two year program and qualified as Psychiatric Nurses. The Motivation Index, based on the TAT administered

in their first year, was unsuccessful in predicting second year final overall averages or scores from the provincial comprehensive examinations of these 17 students.

There were significant correlations between the motivation scores and secondary measures of academic achievement of the second year class (see Table 8). The one individual subject of instruction involved was Psychiatric Nursing II. All three sources of motivation scores (the first scorer, second scorer and average of the two scorers) were intercorrelated at significant levels ($p < .05$) with final student averages in Psychiatric Nursing II. In addition, the Motivation Indices, assigned by the second scorer, were significantly related to scores on the provincial comprehensive examination in Medical-Surgical Nursing.

Interscorer Reliability

The reliability of scoring of the TAT protocols was calculated in two ways: (1) the extent to which the two scorers agreed in the assignment of positive or negative scores was calculated on the basis of percentage of agreement, and (2) the agreement between the Motivation Indices assigned to subjects by one scorer and those assigned by the other was calculated using the Pearson product moment correlation coefficient.

In the scoring of the protocols of the first year class, there was 66% agreement in the assignment of positive or negative scores (172 of 261 stories). In the scoring of the TAT stories of the second year class, there was 67% plus-minus

agreement between the two scorers (164 of 245 stories).

Interscorer reliability for the scoring of the TAT's of the first year class (N = 22) was based on the correlation between the motivation scores assigned by the first scorer and those assigned by the second scorer (see Table 9). There was significant agreement between these two sources of motivation scores assigned to the first year class, $r(20) = .50$, $p = .009$.

For the second year class (N = 21), the interscorer reliability coefficient indicated a significant and adequate level of agreement between the two scorers. The correlation between the Motivation Indexes assigned by the first scorer and those assigned by the second scorer was greater than the .70 level of adequacy, $r(19) = .71$, $p < .001$.

The Characteristics of the Grades

Mean and standard deviations of the achievement measures are displayed in Table 10 and Table 11 for the first year class and second year class, respectively.

Variability in grades. The variability in grades was low in both classes. Low variability reduces the probability of discovering a large correlation between grades and the motivation measure. Kolstoe (1973, p. 194) states: "The size of the correlation coefficient is directly related to the range of scores in both the X and Y distribution."

In the final grades assigned to the first year class, there were no grades below 50%. In the sample of 24 students

Table 9

Motivation Scores from Story Sequence
Analysis of the TAT: Descriptive
and Correlational Statistics

Source of Score	Descriptive Statistics	
	Mean	Standard Deviation
First Year Class:		
First Scorer	94.6	34.6
Second Scorer	131.4	38.2
Average Score	113.0	31.5
Second Year Class:		
First Scorer	77.4	41.2
Second Scorer	116.2	38.9
Average Score	96.8	37.1
Source of Score	Correlational Statistics	
	Source of Score	
	Second Scorer	Average Score
First Year Class:		
First Scorer	.50*	.85**
Second Scorer		.88**
Second Year Class		
First Scorer	.71**	.93**
Second Scorer		.92**

* $p < .01$

** $p < .001$

Table 10
Means and Standard Deviations for Achievement
Measures of the First Year Class

Achievement Measure	Mean	Standard Deviation
Full Class (N = 22)		
Final Overall Average	76.4	6.6
Anatomy and Physiology	74.6	8.4
Nursing I	75.0	8.4
Pharmacology	73.5	8.5
Psychiatric Nursing	73.1	5.5
Intro. Psychology	78.4	7.0
Intro. Sociology	76.6	5.2
Intro. Psychopathology	79.4	10.1
Intro. Family	79.5	7.6
Follow-up Group (N = 17)		
Second Year		
Final Overall Average	76.3	5.4
Provincial Examinations:		
Medical-Surgical Nursing	72.2	6.7
Psychiatric Nursing	69.2	3.9
Social Sciences	72.9	5.6

Table 11
Means and Standard Deviations for Achievement
Measures of the Second Year Class

Achievement Measure	Mean	Standard Deviation
Final Overall Average	73.2	4.0
Medical-Surgical Nursing	71.2	5.5
Psychiatric Nursing II	77.1	4.1
Family Dynamics	71.9	7.9
Psychopathology	69.2	4.6
Psychotherapy	76.4	4.6
Provincial Examinations:		
Medical-Surgical Nursing	78.9	8.7
Psychiatric Nursing	65.3	6.1
Social Sciences	76.2	5.3

Note: N = 21

(each assigned final grades in seven subjects for a total of 168 grades) only three individual grades were below 60%. The standard deviations also suggest low variability. With an average standard deviation for the subjects of instruction of 7.5, 66% of the grades can be expected to be found within a range of 15 marks. The average range of grades assigned to the first year class in individual subjects was 26. That indicates that, on the average, 100% of the grades fell within a 26 point range.

The variability of grades assigned the second year class was less than that of the first year class. The average range of grades assigned in the individual subjects 19 and the the average standard deviation was 5.5.

The variability in the final overall averages assigned the first year class is indicated by a range of 23 and standard deviation of 6.6. The variability in the final overall averages assigned the second year class is indicated by a range of 15 and a standard deviation of 4.0.

Reliability of grades. The reliability of the grades was estimated by correlating final averages assigned in each individual subject of instruction with the grades in each other individual subject of instruction. The arithmetic mean of the correlations, in this correlation matrix, was taken as a measure of the reliability of the grades. Using this measure, the reliability of the grades was .77 and .54 for the first and second year class, respectively.

Time Involved in Providing SSA Scores

When estimating the value of a test instrument, psychologists find it advisable to consider information yield in relation to time input. Records were kept of the time involved in administration of the TAT, scoring the protocols and the training of scorers. These estimates provide an indication of the cost of providing SSA scores.

Administration time. Group administration of the TAT required 2½ hours per class.

Scoring time. Scoring time decreased with experience. An average of 90 minutes was required for the first 20 protocols scored. Those scored after that required an average of 60 minutes per protocol. Since peak concentration was required for scoring, only one protocol could be completed at any one time.

Training time. The initial training project provided one estimate of training time required to teach adequate levels of interscorer reliability. A level of 72% plus-minus agreement was reached by two trainees after 36 hours of training time.

The training of the second scorer was undertaken in smaller steps using criterion scores from the initial training project as a standard. An interscorer agreement level of 70% was reached after 36 hours of training.

Agreement Levels with Expert Scorers

Because of the materials used in the training of scorers it was possible to estimate agreement levels between scorers, in this investigation, and scorers trained and supervised by Dr. Arnold. The training protocols were sample protocols used in previous Story Sequence Analysis research. The scores assigned by the previous researchers were available, in most cases, with these protocols. Since the scores provided by the previous researchers were not used in the training per se, they were available afterwards for assessment of agreement levels with expert scorers. The level of agreement between the first and second scorer and previous researchers, on a plus-minus basis, was 64% and 75%, respectively.

Summary of Major Findings

Tables 7 and 8 indicate that there were some significant relationships between the SSA Motivation Index and final grades. However, the correlations between the SSA scores and the final overall averages were not significant at the $p = .05$ level. The null hypothesis was accepted for this relationship of primary interest.

Interscorer reliability levels when measured by the Pearson product moment correlation coefficient were .50 and .71 for the scoring of the first and second class, respectively. Based on plus-minus agreement, interscorer agreement was 66% and 67% for the first and second class, respectively. Agreement with previous researchers in the scoring of the training protocols, was 64% and 75%.

Chapter V

Summary, Discussion and Conclusions

This study was designed to assess Arnold's Story Sequence Analysis method of scoring the Thematic Apperception Test. The research used protocols from 43 psychiatric nursing students, 22 in the first year of training and 21 in the second year. SSA scores were used in the prediction of the academic achievement of these nurses. In order to assess the complete objectivity of the SSA scoring system the scorers learned this projective methodology from the book (Arnold, 1962). In previous studies training directly involved Dr. Arnold, or another qualified scorer.

Rationale for this Study

One reason for interest in Story Sequence Analysis is that motivation and the measurement of motivation is involved. There is empirical evidence (Brown, 1967) that Arnold's methodology offers the best lead for the further development of a single, more adequate empirical measure of achievement motivation. Theoretical constructs, that can be derived from the SSA scoring system, have been found to be more comprehensive and adequate, in accounting for sources of achievement motivation, than the constructs derived, in a similar fashion, from McClelland's theory (Brown, 1967, p. 145).

Further grounds, for researching the SSA procedure, is to be found in its viability for psychometric assessment and diagnosis. SSA scores had 100% success in discriminating better than average executives from the average executive (Steggert, 1961). SSA scores had 100% success, or nearly so, in discriminating highly rated teachers from low-rated teachers (Burkard, 1958). In three studies (Arnold, 1962; Dulin, 1968; Garvin, 1960), where intelligence measures were combined with SSA scores, correlations with school and college grades were .84 or higher. This empirical data suggests a level of validity of SSA scores that would justify their use for psychometric assessment. Brown (1979) reports from 12 years of selecting candidates for the Jesuit Order that this method is always helpful and, in fact, the best way he found to get insight into the dynamic factors of personality.

The usefulness of Story Sequence Analysis in the therapeutic setting has also been confirmed. Smith (1971) and Ramirez (1970) were able to show positive change in SSA scores of clients in therapy. Those not in therapy showed either no change or a decrease in Motivation Index. In a comparison of projective and questionnaire measures of therapeutic outcome, Smith found SSA scores more satisfactory than scales from the MMPI or 16PF. Story Sequence Analysis has always been the most helpful way of presenting strengths and weaknesses to selection candidates, according to Brown's report (1979). The TAT has been used as a therapeutic tool in the

counselling process (Harrison, 1965) and these reports indicate that Story Sequence Analysis could play a valuable part in that role.

SSA scores are derived from the Thematic Apperception Test, and the fact that a projective methodology has been so highly predictive of achievement goes contrary to the findings of the reviewers (Anastasi, Mischel, 1968; Vernon, 1964). The paradox of projective testing has been the disparity between their widespread use, on the one hand, and their equivocal validity status on the other (Shneidman, 1965, p. 502). Since scientific standards are expected of psychologists both by themselves and others, this reversal in the trend of results with projectives is worth following up. The use of projectives continues and, in fact, a good proportion of employers, 84%, require skill in the use of projective techniques from the clinical psychologist (Levy & Fox, 1975). In order to provide psychologists with an objective scoring system for the TAT, research and development should be continued in Arnold's methodology.

Results

In this investigation some significant correlations between the SSA Motivation Index and final grades in individual subjects of instruction were found ($p < .05$). However, the correlations between the SSA scores and final overall averages were not significant (at $p = .05$). The null hypothesis was accepted for this relationship of major

interest.

For the students in their first year at the time of TAT administration (N = 22), correlations between SSA scores and final averages were from .21 to .24 (not significant at $p = .05$). Correlations between SSA scores of these students and their final grades in the subjects of Pharmacology and Introductory Psychology were significant ($p < .05$). The significant correlations for the first year class (Table 7) were in the magnitude of .36 to .37.

There were 21 students in the second year class. In the relationship of major interest, between SSA scores and final overall averages, correlations were from .21 to .24 (not significant at $p = .05$). However, significant correlations ($p < .05$), ranging from .36 to .44, were found between SSA scores and final grades in Psychiatric Nursing II and between SSA scores and comprehensive examination marks in Medical-Surgical Nursing (Table 8).

Interscorer reliability, calculated by Pearson correlations between the SSA scores assigned by first scorer and those assigned by the second scorer, were .50 and .71, for the first and second year class, respectively. It may be relevant to note that the TAT protocols of the first year class were scored first so that the two scorers had more experience when they scored the second year class.

Interscorer agreement, based on positive or negative scoring of the individual stories, was 66% and 67%, for the

first and second year class, respectively. These agreement levels were almost equivalent with 64% and 75% plus-minus agreement between the SSA scores of the two scorers from this study and the SSA scores of previous researchers. This equivalence of plus-minus agreement levels contrasts with the reliability figures from Brown's (1967) study in which local reliability coefficients greatly exceeded external reliability coefficients.

The attenuation of the correlations between grades and motivation scores by the unreliability of these measures was calculated. The attenuation factor is the product of the square root of the reliability of each measure (Kostoe, 1973, p. 195). This factor was estimated at .62 for both the first class ($.77 \times .50$) and the second class ($.54 \times .71$).

Discussion

Although some correlations between SSA scores and final grades were significant ($p < .05$) there is a large disparity between the magnitude of the correlations found in this study and those reported by some of the previous researchers (Arnold, 1962; Dulin, 1968; Garvin, 1960). The effect of unreliability on validity accounts for only part of this disparity. Other factors that may be responsible for a large part of this discrepancy are: 1) intellectual variables, and 2) contamination of the administration.

The one previous researcher who failed to find a significant relationship between GPA and SSA scores was Brown

(1967). Brown provided some evidence that an extremely difficult math exam from an unpopular teacher, shortly before the TAT administration, induced feelings of anger, frustration, and a sense of failure in his subjects which contaminated the TAT stories. A similar event occurred in this study. The students in this study were addressed by the Director of Nursing, on the day of the TAT administration, and told that the hospital could not guarantee jobs to the graduates of the School of Nursing. The turmoil caused by this major announcement could have contaminated the TAT administration in this study and reduced the validity of the SSA scores.

Intellectual variables may be effecting the validity of SSA scores in two ways. First of all it may be that SSA scores are not as effective for the prediction of the scholastic achievement of students in the lower half of intelligence and achievement as they are in the prediction of achievement of students who are in the upper levels of intelligence and achievement. In three of four correlational studies where information is available (Brown, 1967; Dulin, 1968; Garvin, 1960) it is clear that the students were above average in intelligence. In the earlier studies of high school students (Brown, 1953; McCandlish, 1958; Snider, 1954) the students were selected for admittance to the high school on the basis of achievement tests. In one study (Brown, 1954) only students with an IQ greater than 120 were used as subjects. Youngg (1972) found that academic achievement prediction using

SSA scores was successful for a group of high achievers but not successful for a group of average students. Further study is required to determine whether SSA scores are as successful in predicting the school grades of low achievers as they are in predicting the grades of high achievers.

The large range in academic achievement of the students in the School of Nursing means that a large proportion of variance in grades is due to differences in scholastic attainment and aptitude. The size of a correlation coefficient is directly related to the range of scores in the variables. Seven students were admitted to the psychiatric nursing program without high school diplomas; others had college credits. These students were not screened on the basis of achievement or intelligence tests. These facts suggest that intellectual factors accounted for a much larger proportion of the variance in the final grades of the student nurses than would have been accounted for by this factor in previous studies where selection of the students by intellectual factors was much more rigorous. It also follows that where intellectual variables are more patent than motivational variables are less so. Future studies should use measurements of intellectual variables, the Motivation Index, and other predictor variables so as to assess the relationship of these factors with different populations.

Conclusions

Evidence from this study suggests that new learners of Story Sequence Analysis can produce SSA scores with adequate ($r < .70$) levels of interscorer reliability. Further, there is evidence of adequate levels of agreement between the SSA scores of new learners of this method and experienced scorers. These findings suggest that SSA scores can be learned from the manual with adequate levels of reliability.

Adequate levels of interscorer reliability, however, were not produced consistently in this study. Lack of consistency in reliability coefficients has not been isolated to this study. Brown (1967), Ramirez (1970) and Youngg (1972) also report reliability coefficients that are less than adequate along with coefficients that are adequate. These findings lead to the conclusion that present methods of training scorers using the published scoring system (Arnold, 1962) do not consistently produce adequate levels of interscorer reliability.

It appears that Story Sequence Analysis is not simply an actuarial procedure. It is also a clinical skill that requires prolonged training and supervision. Learning Arnold's SSA system is more difficult than might be expected from Arnold's statement "The method . . . proved comparatively easy to teach and easy to learn" (Arnold, 1962, p. v). Feld and Smith (1958) found that 12 hours of training in Need for Achievement (N Ach) scoring resulted in reliability coefficients

acceptable for research purposes ($r < .80$). Training in this study involved at least 36 hours without achieving equivalent reliability coefficients. It appears likely that other SSA researchers have invested as much or more time in the training of scorers. In his dissertation Brown reports "Learning the Arnold's SSA system and establishing competent interscorer reliability are much more difficult than the training procedure for the other two coding systems [Need for Achievement and Need for Academic Achievement]" (1967, p. 52). Clearly, the length of time involved in training scorers and the lack of consistently adequate reliability coefficients, are drawbacks to the Story Sequence Analysis method of scoring the TAT. Before this method can be said to be fully viable in the applied setting these problems must be resolved.

Limitations of this Study

This research is limited by the fact that it concentrated on a single measure - the Motivation Index derived from Story Sequence Analysis of the TAT. It is clear from the previous discussion that a measure of intelligence would have added to our understanding of the combined role of SSA scores and intelligence in academic achievement prediction.

The use of only two scorers was a weakness of the study. The use of two scorers provided one estimate of interscorer reliability and two estimates of the validity of SSA scores. More estimates would increase the confidence that one can

place in the results and increase understanding of the effect of the individual scorer on the results.

Implications for Further Research

The primary focus of future research should be the development of a standardized training procedure. This procedure should be based on the use of the published scoring system, perhaps revised and simplified, and specially compiled sets of practice materials. Training should develop high levels of agreement between new learners of this method and scoring standards derived from experienced SSA scorers. As Feld and Smith (1958) found, a written explanation of the expert's scoring would be a necessary component of the training materials. Ideally principles of programmed learning would be used in devising the training procedure.

The use of consensus scores (those agreed on by two or more scorers) could be used by future researchers. One possible advantage of consensus scores is their greater stability and validity. Also, the consensus scoring procedure, applied every five protocols, for example, allows for ongoing training among the scorers. The researcher should record both the scores of the individual scorers, before conferencing them, and the final consensus scorers. This procedure will allow the researcher to compare the validity of consensus scores to scorers derived from the individual scores.

Future investigation might also:

- 1) investigate the effect of experimentally induced

frustration on SSA scores.

2) use SSA scores as part of a battery of tests and other indicators in order to develop a model for the prediction of academic achievement.

3) follow-up research on the effectiveness of SSA scores in differentiating between effective and ineffective teachers.

4) study changes in motivation levels of clients, students or employees in relation to the motivation levels of counsellors, teachers and supervisors.

5) determine whether changes in motivation level, as reflected in SSA scores, correspond with behavior and performance changes (for example, is an increase in the Motivation Index of a student associated with increased G.P.A.?).

It appears that Story Sequence Analysis is not simply an actuarial procedure that can be easily learned from Arnold's book (1962) and applied to scoring the TAT. Considerable time and effort are required in learning the scoring procedure and in scoring the individual sets of TAT stories. This cost should not be allowed to become a deterrant if, as evidence suggests, Arnold's methodology offers the best lead to the development of a practical technique for the measurement of motivation and prediction of achievement.

REFERENCES

- Abelson, R. P. & Tukey, J. W. Efficient conversion of non-metric information into metric information. In E. R. Tufte (ed.), The quantitative analysis of social problems. Reading, Mass.: Addison Wesley, 1970.
- Anastasi, A. Psychological testing (3rd Ed.). New York: MacMillan, 1968.
- Arnold, M. B. A demonstration analysis of the Thematic Apperception Test in a clinical setting. Journal of Abnormal Social Psychology, 1949, 44, 97-111.
- Arnold, M. B. Story sequence analysis: A new method of measuring motivation and predicting achievement. New York: Columbia University Press, 1962.
- Behrens, L. T. & Vernon, P. E. Personality correlates of over-achievement and under-achievement. British Journal of Educational Psychology, 1978, 48, 290-297.
- Brown, J. E. Personality dynamics of high and low academic achievers in high school; a modified Thematic Apperception Test reliability study. Unpublished master's thesis, St. Louis University, 1953.
- Brown, J. E. Motivational factors in college achievement (doctoral dissertation, Catholic University of America, 1967). University Microfilms No. 67-17, 126.
- Brown, J. E. Personal communication, October, 1979.
- Brown, W. R. & McGuire, J. M. Current assessment practices. Professional Psychology, 1976, 7, 475-84.
- Burkard, M. I. Characteristic differences, determined by TAT sequential analysis, between teachers rated by their pupils at the extremes of teacher efficiency. Unpublished doctoral dissertation, Loyola University, Chicago, 1958.
- Burkard, M. I. Discernment of teacher characteristics by TAT sequence analysis. Journal of Educational Psychology, 1962, 53, 279-287.
- Dulin, J. T. Prediction of academic achievement in college. Unpublished doctoral dissertation, Loyola University, Chicago, 1968.

- Feld, S. & Smith, C. P. An evaluation of the objectivity of the method of content analysis. In J. W. Atkinson (ed.), Motives in fantasy, action and society. Princeton: Van Nostrand, 1958.
- Fields, R. M. Motivational factors of unwed mothers. Unpublished master's thesis, Loyola University, Chicago, 1964.
- Garvin, J. A. A Thematic Apperception Test study of non-intellective factors related to academic success on the college level. Unpublished doctoral dissertation, Loyola University, Chicago, 1960.
- Harrison, R. Thematic apperceptive methods. In B. B. Wolman (ed.), Handbook of clinical psychology. New York: McGraw-Hill, 1965.
- Honor, S. H. TAT and direct methods of obtaining educational attitudes of high and low achieving high school boys. (Doctoral dissertation, Hofstra University, 1971). University Microfilms No. 71-72, 267.
- Klopfer, W. G. & Taulbee, E. S. Projective tests. In M. R. Rosenzweig & L. W. Porter (eds.), Annual Review of Psychology. Palo Alto: Annual Reviews Inc., 1976.
- Kolstoe, R. H. Introduction to statistics for the behavioral sciences. Homewood, Illinois: Dorsey, 1973.
- Labovitz, S. The assignment of numbers to rank order categories. American Sociological Review, 1970, 35, 515-524.
- Lazarus, R. S. A substitutive-defensive conception of apperceptive fantasy. In J. Kagan and G. S. Lessen, eds. Contemporary issues in thematic apperceptive methods. Springfield: Thomas, 1961.
- Levy, M. R. & Fox, H. M. Psychological testing is alive and well. Professional Psychology, 1975, 6, 420-434.
- McCandlish, L. A. An investigation of a new method of T.A.T. analysis by a prediction study of high and low academic achievers. Unpublished doctoral dissertation, Loyola University, Chicago, 1958.
- Mischel, W. Personality and assessment. New York: Wiley, 1968.
- Murray, H. A. Thematic apperception test (Manual). Harvard College, 1943.
- Murstein, B. I. Handbook of projective techniques. New York: Basic Books, 1965.

- Najjar, B. Differences in motivational patterns of inadequate and inept psychopathic deviant prison inmates as revealed by TAT story sequence analysis. Unpublished doctoral dissertation, Loyola University, Chicago, 1967. Cited in Ramirez, 1970.
- Nie, H. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. & Bent, D. H. SPSS: Statistical package for the social sciences (2nd ed.). New York: McGraw Hill, 1975.
- Petrauskas, F. B. A TAT and picture-frustration study of naval offenders and non-offenders. Unpublished doctoral dissertation, Loyola University, Chicago, 1958.
- Quinn, T. A discriminative scoring system for the TAT: An investigation. Unpublished master's thesis, Loyola University, Chicago, 1959. Cited in Garvin, 1960.
- Quinn, T. Difference in motivational patterns of college student brothers as revealed in the TAT, the ratings of their peers and the ratings of their superiors: a validation study. Unpublished doctoral dissertation, Loyola University, Chicago, 1961. Cited in Arnold, 1962.
- Ramirez, J. A. Changes in motivation during client-centered therapy as measured by the story sequence analysis of the TAT. Unpublished doctoral dissertation, Loyola University, Chicago, 1970.
- Smith, B. An evaluation of group therapy using the 16 PF, M.M.P.I., and TAT. Unpublished master's thesis, Loyola University, Chicago, 1971.
- Snider, L. B. Personality differences between high and low academic achievers in high school: A Rorschach and Thematic Apperception Test Study. Unpublished doctoral dissertation, Loyola University, Chicago, 1954.
- Steggert, F. X. An analysis of some personal and executive characteristics of participants in a university program of executive development for federal personnel. Unpublished doctoral dissertation, Loyola University, Chicago, 1961. Cited in Arnold, 1962.
- Tyrell, D. J. When love is lost. Waco, Texas: Word Press, 1971.
- Vernon, P. E. Personality assessment: A critical survey. London: Methuen, 1964.
- Youngg, J. H. The fakability of the TAT under conditions of instruction to fake and test clues: Arnold's story sequence analysis for achievement motivation. Unpublished doctoral dissertation, Hofstra University, 1972.

APPENDIX

Main Headings in Story Sequence Analysis
Scoring SystemAchievement Category, Headings 1 - 6

- 1 Goals, purposes, motives, values.
- 2 Means taken toward goal. Includes: attitude toward work recreation and rest.
- 3 Adaptability of goals and means.
- 4 Influence of others on achievement. Includes: help, advice, legitimate commands, approval and choosing one's life course.
- 5 Overcoming failure.*
- 6 Attitudes toward success.*

Right and Wrong Category, Headings 7 to 9

- 7 Actions. Includes: wrongdoing, injustice, duty, accidentally harmful action.
- 8 Intentions. Includes: ill intentions, negative emotions, wrong attitudes, temptation and revenge.
- 9 Punishment. Includes: unjust punishment and desirable and undesirable consequences of just punishment.

Human Relationships Category, Headings 10 to 14

- 10 Good relationships, love, friendship, marriage. Includes: durability of relationships, the expression of good relationships, love and one's principles.

- 11 Bad relationships, hate, bitterness, long lasting quarrels.
Includes: causes, effects, venting feelings, correcting relationships and separation.
- 12 Influence of others on one's action choice. Includes:
help, reasonable advice, commands, bad influence, illegitimate pressure, reasonable independence and blind dependence.
- 13 Influence on others. Includes: positive and negative influence, helping and manipulating.
- 14 Attitudes towards people, things, God, nature, life.
Includes: optimism and pessimism.

Reaction to Adversity Category, Heading 15

- 15 Adversity, loss, separation, disappointment, harm, danger and terror. Includes: overcoming adversity by action and attitude, undesirable reactions to and effects of adversity.

* These headings have been simplified by this author from Arnold's categories "Consequences of success (failure)" and "Attitudes connected with success (failure)".