AN ABSTRACT OF THE THESIS OF

Cindy Katherine Primeau for the Master of Science in Clinical Psychology presented December, 1989

Title Relationship Between Scores on the Gorham Proverbs Test and the Arlin Test of Formal Reasoning

Abstract Approved: [signature]

The purpose of this study was to investigate the relationships between the Proverbs Test (PT) and the Arlin Test of Formal Reasoning (ATFR) and its eight subtests. Additional issues explored were the percentage of college students in a small midwestern university performing at formal operational levels and whether there were significant differences in performance on the two tests by males and females. A sample of 56 subjects (17 males and 39 females) between the ages of 18-20 was drawn from four sections of a beginning level psychology course. Both tests were administered to each subject. The scores obtained on the Proverbs Test, the ATFR and its eight subtests were recorded for each subject, as were each subject's age and gender. Pearson product-moment correlational coefficients were calculated to determine the relationships between the ATFR Total score, the eight ATFR Subtest scores, and the Proverbs Test Abstract, Concrete and Adjusted scores. In addition, *t*-tests

were computed to determine any significant differences in performance on the two tests by males and females. Subjects were categorized into five cognitive levels as assessed by the ATFR and the percentage of students at each level was determined. Also calculated was the percentage of college students who were performing at the formal operational level.

The relationships between the ATFR Total score and the Proverbs Test Abstract and Adjusted scores were found to be significant. The relationship between the Proverbs Test Concrete score and the ATFR Total score was a significant negative one. The Proverbs Test Abstract score correlated significantly with scores on the ATFR Subtests II, Correlations; IV, Combinational Reasoning; and VI, Forms of Conservation. The Proverbs Adjusted score also correlated significantly with the above ATFR Subtest scores with the addition of Subtest III, Probability. A significant negative correlation was found between the Proverbs Test Concrete score and and the score on the ATFR Subtest III, Probability.

No scores on the Proverbs Test correlated significantly with scores on the ATFR Subtests I, Multiplicative Compensations; VII, Mechanical Equilibrium; or VIII, Frames of Reference. Significant

differences were found between male and female group means on the ATFR Total score as well as scores on the ATFR Subtests I, Multiplicative Compensations; V, Proportional Reasoning; and VIII, Frames of Reference. There were no significant differences between male and female group means on the Proverbs Test.

The small population, homogeneity of the data, and the limited age and intellectual range call for further study to generalize results. It is suggested, therefore, that further research is needed.

Relationship Between Scores on the Gorham Proverbs Test and the Arlin Test of Formal Reasoning

---

A Thesis

Presented to

the Division of Psychology and Special Education

Emporia State University

In Partial Fulfillment

of the Requirements for the Degree

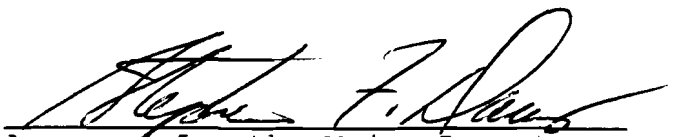
Master of Science

---

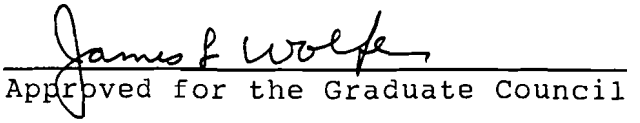
by

Cindy Katherine Primeau

December, 1989



Approved for the Major Department



Approved for the Graduate Council

ACKNOWLEDGEMENTS

My greatest respect and sincerest gratitude are paid to Dr. David Dungan, Dr. Loren Tompkins, and Mr. Howard Carvajal. Their support and guidance in this endeavor made it possible for this study to become a reality. Their kindness will not be forgotten.

To my husband, Dave, goes my deepest appreciation for his patience, love, and understanding. Thank you for being there.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

APPENDIX PAGE

CHAPTER 1

Introduction

Formal operational thought as described by Piaget (Inhelder & Piaget, 1958) is the culmination of the process of intellectual development begun in childhood. The work of Jean Piaget has been an attempt to isolate and describe the mental structures upon which reasoning processes are based. His overall aim was to trace the development of intelligence as it comes to deal with increasingly complex problems or as it deals with simple problems in increasingly complex ways.

Piaget describes four major stages of intellectual growth. Recent research and the attention of this study are focussed on assessment of the fourth stage of formal operations.

The formal operational stage, beginning somewhere between 12-15 years of age and continuing through adulthood, is characterized by the development of the ability to use hypothetical reasoning based on a logic of all possible combinations of variables and the ability to perform controlled experiments. Formal thinking goes beyond the manipulation of empirically given data, as in the concrete operational thought of children, to generalizing from these data by way of

isolating variables and deducing potential relationships which can later be verified by experiment. In formal thought, verbal statements take the place of objects, providing the opportunity to operate with propositions (Lovell, 1971).

College and university instructors have become concerned with the assessment and stimulation of the formal reasoning ability of their students. Whereas Inhelder and Piaget (1958) propose that the acquisition of formal operations is universal for normal adolescents (provided their social culture and education are conducive, Piaget, 1972) several studies report 50% or less of the adolescents and adults tested demonstrated this ability (McKinnon & Renner, 1971; Ross, 1974; Schwebel, 1975). Science teachers at the high school and college level are concerned that their students are not capable of or are not demonstrating the type of reasoning necessary to deal with a science curriculum (McKinnon & Renner, 1971). The ability to use formal operational thought is necessary across major content areas such as history (Lovell, 1971), numerical and verbal analogies (Lunger, 1965 as cited in Lovell, 1971), and literature (Hardy-Brown, 1979) as well as in the sciences.

The determination of adult levels of cognitive functioning (concrete, transitional or formal) has in the past been a laborious undertaking. The evaluation of formal reasoning capabilities by Piaget's clinical method (Inhelder & Piaget, 1958) involves the use of bulky laboratory materials, requires the expertise of a trained evaluator and is time consuming. The development and validation of group administered tests and group paper-and-pencil tests to replace the individual interview technique of the clinical method have been the concern of educators and researchers for more than a decade (Lawson, 1985).

The need for a way to investigate the construct of formal thinking required researchers to be able to assess large groups of individuals with reliable and valid measures in order to classify them as to the various stages of cognitive development. With this accomplished, instructional methods and materials could then be developed appropriate to each stage (Ahlawat & Billeh, 1982).

Review of the Literature

Many group tests of formal reasoning combine the precision of conventional psychometric tests with the measurement of abilities described by Piaget's original interview method (Shayer, Adey & Wylam, 1981). Efforts

to develop useful group measures of Piagetian cognitive levels fall into three major categories: (a) science content based instruments which pair demonstration with some form of written answers; (b) written science content based instruments; and (c) written instuments which use non-science terminology. Research on the three types of measures occurred concurrently from the early 1970s to the present.

Group tests that are easily administered, readily scored and still retain as many as possible of the attributes of the original Piagetian tasks have been the goal of science educators. Rowell and Hoffman (1975) attempted to translate two Piagetian-type problem situations into suitable forms for group administration. Inhelder and Piaget's (1958) chemical change experiment and the pendulum experiment were chosen due to their structural similarity (but which involved very different subject matter) and the use of unsophisticated apparatus used to complete the test. Subjects were given a set of written instructions and worksheets as well as apparatus with which to complete the problems. The instructor prepared the subjects by reading through all directions on the written instruction sheet with them and then allowed the subjects to manipulate experimental apparatus as they

solved the problems. The sample was chosen from eight classes from the first four years of a South Australian metropolitan high school (age range 12-16 years). Results of the study indicated that the two tasks provided relatively the same measure of developmental level (product-moment correlational coefficient, r = .56), and that it was possible to translate Piagetian indicators of developmental level into a group administered form. However, this method of testing for cognitive developmental level was not compared with any other measure such as the clinical interview to help determine its validity. Further difficulty arises with the need for large quantities of laboratory equipment. This and the time needed for testing restrict assessment of large groups.

Lawson (1978) felt that a test of formal reasoning needed to relate as closely as possible to the Piagetian individual clinical interview method to retain the "motivating aspects and sense of meaningfulness that arise from physical materials and equipment" (p. 11) while retaining the ease of administration and scoring and the use of a limited amount of laboratory equipment. The format of his Classroom Test of Formal Reasoning (CTFR) involved one set of materials that were used by the instructor to demonstrate situations and pose

questions to the class as a whole. Students then responded individually in writing on their own test booklets.

Fifteen formal operational items were compiled from Inhelder and Piaget's (1958) work as well as from various other authors. Elements included operations used in the isolation and control of variables, combinatorial reasoning, probabilistic reasoning, and proportional reasoning. The test booklets contained questions pertaining to demonstrations in each area with a number of possible answers. Students were instructed to choose the best answer and then explain in writing their choice.

Subjects for the study included 513 students in grades 8-12 in two middle to upper-middle class suburban communities in the San Francisco Bay area. Classes of students were selected from the entire range of student abilities across all three grades. The number of male and female students was nearly equal with ages ranging from 12-18 years.

Testing time of each class was from 75 to 100 minutes in two consecutive periods. In order to test the validity of the group test, 72 subjects were randomly selected and administered a battery of four Piagetian tasks conducted by trained interviewers.

Reliability estimates of these tasks had been established by Lawson, Nordland and DeVito (1974) with test-retest correlation coefficients ranging from .48 to .78. The Kuder-Richardson 20 estimate of reliability in Lawson's study was .78 on the CTFR. Lawson (1978) states that the validity of the Piagetian tasks had been determined by numerous investigators. Lawson reports validity based on three types of evidence: (a) construct validity, (b) relationship between CTFR total scores and the level of subject response on the Piagetian tasks, and (c) a prinicpal components analysis of the CTFR and all four interview tasks. Construct validity was determined by a panel of six judges considered experts in Piagetian research who responded with 100% agreement that the test items appeared to require concrete and/or formal reasoning.

Pearson product-moment correlations between the CTFR and level of response on the two tasks were .75 and .65 ($p < .001$), respectively. Correlations between CTFR and the summed task scores was .76 ($p < .001$).

Principal component analysis based on the 72 cases using the Piagetian tasks yielded tentative results. Lawson hypothesized that since all of the items, with the exception of the conservation of weight, were designed to measure aspects of formal operational

reasoning, the analysis should yield only two principal components: a concrete component and a formal reasoning component. Results indicated this to be true with the unexpected addition of a third factor which was identified as an early formal reasoning factor.

Lawson (1978) cautions that the CTFR tends to place students in lower categories than does the interview method. This may cause the underestimation of the abilities of the classroom as a whole.

A test combining written work with classroom demonstration was developed by Tobin and Capie (1981). Items used on the test and the general format for its administration were taken from the work of Lawson (1978). The Test of Logical Thinking (TOLT) requires students to write an explanation of why they chose a particular multiple choice answer dealing with a particular demonstration. One difficulty with this test was the inability of some students to formulate clear justification for their answers. Because of this initial problem, the TOLT was modified to provide multiple choice justification statements. The revised test also included videotaped demonstrations of administration procedures. Reliability and validity studies were conducted on a sample of 353 students in middle school grades 6, 7, and 8, and on a sample of 82 physics and

chemistry students from grades 11 and 12. College students in science courses also participated (N = 247). The sample was not chosen to be representative of grade 11 and 12 students, but to investigate whether physics and chemistry students would include a larger proportion of formal thinkers due to the nature of Piagetian tasks dealing mainly with science content. Results indicated that there was a gradual increase in performance from grades 6 through college. Tasks were moderately intercorrelated (r = .35-.53). Factor analysis suggested a common structure underlying performance on each item. Correlation between performance on Piagetian clinical interviews given to a sample of 25 college students and 63 high school students and their scores on the TOLT was .80. The authors recommend further testing of the TOLT with large samples that encompass the entire range of formal reasoning ability.

Shayer, Adey, and Wylam (1981) discuss a set of Science Reasoning Tasks (SRTs) first developed in 1979. These group tests assess concrete and formal reasoning strategies much as Lawson (1978) did with his Classroom Test of Formal Reasoning. Each SRT concentrates on one problem, rather than asking the student to solve a selection of diverse problem situations. Instructor

demonstration is paired with individual written answers by the students. Validity and reliability data can be found in Shayer and Wharry (1973).

A group administered test of formal thought was developed in 1979 by Staver and Gabel. The Piagetian Logical Operation Test (PLOT) combines videotape presentation such as that of the TOLT (Tobin & Capie, 1981) with an objective multiple choice written test booklet. PLOT questions are similar to those asked in clinical interviews. The test can be administered to classroom size groups and is easily scored. The authors examined construct validity via multiple methods and traits. Methods included the PLOT, the Piagetian clinical interview, the Lorge-Thorndike Intelligence Test (Form 1, Levels C, D, E), and the Cognitive Abilities Test (Form 1, Level G). Traits measured were conservation of volume by liquid displacement, separation and control of variables, combinatorial analysis, proportional thought, and verbal, non-verbal, and quantitative abilities. The Lorge-Thorndike Intelligence Test and the Cognitive Abilities Test are measures of general mental ability. The sample of 126 subjects from grades 10-12 in a south central Indiana consolidated school corporation contained an equal number of males and females for each grade.

Results indicated, by the correlation of methods and traits, that evidence only for convergence between Piagetian measures exists while little evidence of dicriminant validity exists for the Piagetian and general intelligence tests. The factor analysis, however, provided clear evidence for discriminance, but little support for convergence. The correlation between the PLOT total score and the total interview score (r = .59) also supports convergence. The authors concluded that convergent and discriminant validity were partially established.

No significant differences in performance with respect to gender were shown for PLOT and its scales. Learning effect was present with the clinical method, but absent in PLOT. At this point, the authors recommend further test development, but feel that PLOT is a reliable, valid, efficient, and practical tool to aid in the determination of cognitive level assessment for the purpose of curriculum modification.

It is interesting to note that during the 1970s researchers in other countries were working to develop written instruments to measure concrete and formal thinking. Tisher (1971) discusses a Piagetian Questionnaire used to ascertain levels of cognitive

development in secondary schools in Australia. Criteria from Inhelder and Piaget (1958) were used to develop the questions and administration included the demonstration of each experiment for which multiple choice answers were available. The questionnaire was paired with a battery of Piagetian tasks in order to determine if they measured the same attributes. A group of 232 pupils, 138 males and 94 females, was drawn from eight junior high classes of country high schools in Australia. Out of this sample, 57 were also given the battery of Piagetian tasks. An agreement of 77% in classification was found between the written test and the clinical interviews. No significant sex differences were found in performance on the questionnaire.

Measures of formal operational thinking using Piaget's criteria and certain elements of the clinical interview, such as student manipulation of materials, instructor demonstration, and videotaped demonstration paired with written questions appear to be one step away from his original methods of determining cognitive levels. The next step to a completely written format to measure concrete and formal thinking occurred concurrently with the advent of the mixed format test. Many continued to utilize science content based

questions; however, some moved away from this to questions couched in non-scientific terms.

Two tests which use more traditional science based content in the development of items on a written measure are the Inventory of Piaget's Developmental Tasks (IPDT) discussed in Bender and Milakofsky (1982) and the Tomlinson-Keasey and Campbell's (unpublished) Chemical Puzzle used in a study by Tschopp and Kurdek (1981). The former is a 72 item, untimed, multiple-choice paper-and-pencil inventory. The authors cite a reliability and validity study by Patterson and Milakofsky (1980) stating that the Inventory of Piaget's Developmental Tasks (IPDT) yielded results similar to those of individual interviews in a sample of third, sixth, and ninth graders as well as college freshmen and sophomores. Bender and Milakofsky (1982) examined the relationship between IPDT and Scholastic Aptitude Test (SAT) scores of 225 college students enrolled in two different levels of first year chemistry courses. Results indicated that there was a significant relationship between the mathematics SAT and IPDT scores (r = .55) and a somewhat lower correlation between the verbal SAT and IPDT scores (r = .24). The authors concluded that the IPDT has some predictive value for college performance in chemistry

and that "the correlations of .55 for the mathematics SAT and .24 for the verbal SAT also mean that the IPDT is sufficiently independent of the traditional measures of aptitude for colleges to warrant further investigation" (p. 212).

The written Piagetian Task Instrument (PTI) developed by Walker and Mertens (1979) was based on the Piagetian Task Instrument constructed by Sayre and Ball (1975). The 1975 version used a set of tasks which met the requirements of formal operational thought as specified by Inhelder and Piaget (1958). The administration was similar to the clinical interview method. Six formal tasks were selected from the earlier version of the PTI and translated into written form. These tasks were equally balanced among items requiring propositional logic, combinatorial logic, and hypothetical reasoning. The content of the questions differs from the original tasks, such as where combinatorial logic was assessed by means of flasks of liquid chemicals combined with drops to determine the number of color changes possible; the same task in written form asks for the number of all possible combinations of ice cream cones made with five different flavors. Although using the same reasoning patterns, the

written questions may appear less threatening and more familiar to those who might feel intimidated by science based questions.

Construct validity between the interview results and the written responses was determined by a "logical analysis of the relationship between the problem presented in the task and the reasoning pattern required to solve the problem" (p. 213). Though the data are not given, the authors concluded that the written task did require the reasoning pattern it was designed to detect.

Pilot testing of the written instrument was useful in refining the clarity and readability of the questions. Correctness or incorrectness of each written task was based on the "explain your answer" portion of each question to insure that incorrect mathematical calculations did not reduce scores on the test.

A similar measure (Barnes, 1977) used a written questionnaire form with problems taken from Karplus and Karplus (1970) and "puzzles" taken from unnamed sources. Multiple choice answers were available with some questions requiring further explanation. This study sought to investigate to what extent cognitive levels of students predicted final grades in physics courses. The authors found generally low correlations between grades

in physics and scores on the Piagetian questionnaire. This may indicate that those who do not reason formally may do as well as those who do; however, the authors also concluded that other factors may have been involved, such as the validity of their measure or clarity of the questions.

Questions on written tests of cognitive level are often translated from science and mathematical terminology to more familiar types of word problems. The Island Puzzle by Karplus and Karplus (1970) uses such a format. The test proposes to measure "abstract reasoning ability" by providing clues to a situation represented by the drawn puzzle. The individual must analyze these "clues" and use them to draw certain conclusions. Written answers plus explanations of those answers are then entered on the test sheet. Clues are presented orally by an instructor, but are not available to refer to by the subject. Subjects are then categorized by cognitive level according to the quality of their answers and explanations. The authors found that a surprisingly small percentage of subjects tested reached the formal level--even among college physics teachers. This may have been in part due to the administration procedures.

Blake, Lawson, and Nordland (1976) investigated whether the Islands Puzzle (Karplus & Karplus, 1970) does in fact measure the same psychological parameters as the tasks employed by Inhelder and Piaget (1958). Following administration of the Karplus Islands Puzzle to 126 high school students, each of the students was interviewed individually with a battery of Piagetian tasks. Several different forms of the Islands Puzzle plus the Piagetian tasks were also given to a group of 160 college students. Moderate, but significant correlations exist between the Islands Puzzle and the Piagetian tasks. The range of correlation was from .28 to .47. Blake, et al. recommend that teachers not rely solely on the Islands Puzzle to measure student's cognitive levels.

Raven's Test of Logical Operations (RTLO), while not made up of Piagetian tasks, does apply the problem solving rules described in Inhelder and Piaget (1958). Raven (1973) states: "although it uses the same inference patterns (logical operations) that the Piaget tasks use, it employs totally different goal objects and precepts" (p. 378). The general format of the test items is the presentation of problems in pictorial form followed by a brief question. Pictures of possible solutions are provided from which the subject may choose. The test

was administered to 424 students from grades 3-9 in New York schools. Statistical evidence supports the reliability and validity of this measure.

A more recent addition to the list of pencil-and-paper tests, and the focus of this study, is the Arlin Test of Formal Reasoning (ATFR) (Arlin, 1982). The ATFR, while based on the eight formal concepts as defined by Inhelder and Piaget (1958), represents these concepts in non-science, non-mathematical terms and examples. Arlin (1984b) states: "each concept represents a form of thinking and not necessarily the content of that thinking. They are known by the fact of their varied applications rather than being concepts in the narrow sense of the term" (p. 2).

This objectively scored, multiple choice test provides pictorial examples of each question and asks the subject to choose a justification or explanation for answers to the original problem. In a validity study (Arlin, 1982), the multiple choice test was paired with a battery of Piagetian tasks administered in clinical interview. Subjects were 38 male military recruits from the ages of 17.9 to 19.11 years. Findings indicate a highly significant general level of validity. Arlin (1984b) reports a range of correlations from .80 to .90 between the ATFR and Shayer's Science Reasoning Tasks

(SRTs), valid measures pairing demonstration with a written format.

It appears appropriate to investigate the relationship between this measure of Piagetian cognitive levels and a written test in another content area, such as the Gorham Proverbs Test (PT) (Gorham, 1956a), which proposes to measure abstract (formal) and concrete thinking. Support for such a move comes partially from a study by Hardy-Brown, (1979) who successfully used poetry interpretation to assess levels of thinking from concrete to formal. She states:

> An important indication of formal manipulation of a literary work ought to involve the discrimination of the literal (or narative) from the symbolic level. The ability to perceive and consistently discuss the symbolic level in a literary work requires the creation of hypotheses and propositions not given within the empirical data (i.e., the story or poem), in the same way that a formal approach to a scientific problem begins with hypotheses and probabilities rather than limiting itself to a direct reorganization of the perceived data" (p. 129).

The author hypothesized that responses to the interpretation of a poem could be categorized into three

general groupings. Concrete thinkers might be expected to concentrate on the literal objects or people and on the surface story line in a poem. Transitional (from concrete to formal) thinkers may acknowledge both the literal and symbolic levels, yet not be able to reconcile the two. Fully formal thinkers ought to be able to successfully reorganize and integrate the two levels in their responses. Hardy-Brown (1979) investigated this by way of a pencil-and-paper test of traditional Piagetian tasks (an unpublished test devised by Tomlinson-Keasey and Campbell of the University of Nebraska) and the analysis of a literary task.

Both measures were administered to 30 college sophomores and seniors. An interviewer met with each student to record his/her interpretation of a poem. Chi-square analysis of the operational groupings and performance on the poetry revealed significant associations in most cases. Hardy-Brown (1979) states: "this study demonstrated a means of effectively translating logical operational criteria from one domain to another, thereby making valid comparison of performance between domains possible" (p. 134).

While not a test specifically designed for assessment of levels of cognitive ability as Piaget

describes them, the Gorham Proverbs Test (Gorham, 1956a) does propose to assess abstract and concrete thinking abilities. The Proverbs Test was developed in 1956, with a final printing in 1975, as a screening device to differentiate normals from schizophrenics in a psychiatric population. The norm-referenced test purports to measure the level of abstract verbal functioning and requires the subject to translate the "concrete symbols of a proverb into generalized concepts" (p. 435). The test is a multiple choice, 40 item instrument which can be administered in group form or individually. The group form generally takes less than one half hour to complete and is easily scored with the use of a stencil. While the majority of research on the Proverbs Test has been in the area of psychiatric evaluation (Elmore & Gorham, 1957; Fogel, 1965; Gorham, 1956a; Lasky, 1967; Smith, 1971) several studies have dealt with intellectual functioning and abstract reasoning (Gorham 1956b; Gorham, 1963).

Martin (1967) studied the relationship between scores on the Proverbs Test and scores on an experimental Test of Abstract Reasoning (TAR). The TAR appeared to emphasize reading ability, power of concentration, and the ability to follow exacting instructions using a "street map" concept. The TAR correlated highly enough

with the Proverbs Test (r = .45) to indicate that each seemed to be measuring similar intellectual processes. No sex differences in performance were found on either test.

Jurjevich (1967) found significant correlations between the Proverbs Test, Raven's Progressive Matrices, and the Wechsler Adult Intelligence Scale (WAIS). The Proverbs Test had the highest correlation with the WAIS Verbal Scale. It was determined that the Proverbs Test provided an acceptable estimate of intellectual ability.

Proverbs have been used historically to determine levels of cognitive ability. Piaget (1923) used proverbs with young children to determine the quality of symbolic thought. Children under 11 years of age could assign meaning to the proverb, but only in relation to the concrete symbols; they were not able to generalize to abstract interpretations.

Proverbs were added to the 1955 revision of the Wechsler Adult Intelligence Scale, "because of their reported effectiveness in eliciting paralogical (illogical) and concretistic thinking" (Matarazzo, 1972, p. 201). Richardson and Church (1956) studied proverb comprehension along three dimensions: specific-general, literal-figurative, and physiognomic-articulated. Sixty-four children, ages 7.4-12.5, and 30 adults

interpreted seven common proverbs. Significant changes in interpretation were seen between children and adults. Adults generally provided interpretations of the proverbs as general statements instead of concrete representations, were able to comprehend the metaphorical nature of the proverb and coordinate the two levels of meaning, while the children could not. However, many of the adults also reacted in a concrete manner to the proverbs, while some began with a more primitive interpretation and then were able to progress to a more mature one.

Billow (1975), in phase two of his study, explored the relationship of proportional metaphor and proverb interpretation to formal thought. Thirty subjects with 10 in each of three age groups (9-10, 11-12, and 13-14) were given 12 proverbs to interpret as well as a formal task (combinatorial reasoning) to solve. Results indicated that the relationship between proverb comprehension and the formal operational task was not significant (r = .12). There was a significant rise in the number of proverbs comprehended between the ages of 9 and 11, and 11 and 13 (less than 1 out of 12 solved to 5 out of 12 solved) suggesting that a relationship may exist, but is not tapped by the use of the combinatorial task as the sole measurement of formal operational

performance.

Purpose of this Study

The purpose of this study, then, is to examine the relationship between the Gorham Proverbs Test (PT) and the Arlin Test of Formal Reasoning (ATFR). This examination will include correlation of the two total test scores as well as the subtest scores of both tests.

The Arlin Test of Formal Reasoning is based on the concept of formal operational thought as described by Inhelder and Piaget (1958) and describes eight formal schemes relating to formal thought, rather than only one or two. If a significant relationship can be established between these two measures, the Proverbs Test may be used as a quick screening measure to aid in the determination of college students' ability to reason formally, an ability which may be necessary for learning at the college level (Schwebel, 1975). With identification, those students who do not appear to be operating at a formal level and who may be experiencing difficulty with a college curriculum could be taught in more concrete ways (Renner & Paske, 1977; Schneider & Renner, 1980) or taught to apply formal reasoning strategies (Danner & Day, 1977).

Additional issues to be investigated in this study

include the determination of the percentage of freshman students, ages 18-20, in a small midwestern university who appear to be operating at the formal level and the investigation of possible gender differences in performance on written tests of cognitive levels. As stated earlier, a large percentage of college and university students may not be demonstrating formal levels of thinking (Ross, 1974). Females tend to do more poorly on Piagetian tasks than males (Lawson, 1975; Tomlinson-Keasey, 1972), while written tests appear to elicit relatively the same level of performance from both males and females (Staver & Gabel, 1978; Tisher, 1971). No gender differences are reported in performance on the Proverbs Test (Gorham, 1956a). Information as to this factor is not available for the Arlin Test of Formal Reasoning (Arlin, 1982) due to norming on an all male population. This study will endeavor to shed some light on these issues.

CHAPTER 2

Method

Subjects

Fifty-six subjects, 17 males and 39 females, were administered the Arlin Test of Formal Reasoning and the Proverbs Test. The subjects ranged in age from 18-20 with a mean age of 18.6. The male group had a mean age of 18.5, while the female group had a mean age of 18.6. Subjects were freshmen students from 4 sections of a beginning level psychology course in a small midwestern university who volunteered for this study. Permission was granted by the Human Subject's Committee for the testing of these subjects for this study. Scores from the 56 pairs of tests administered were analyzed to obtain the data for this study.

Prior to testing, consent forms were given to each subject explaining the testing procedures, confidentiality, and the subject's right to withdraw from the study at any time (see Appendix A). The subjects were asked to provide their age and sex on the answer forms as they would later be identified only by these characteristics.

Instruments

The Best Answer Form or multiple choice form of

the Proverbs test is designed to measure the level of abstract verbal functioning of adolescents and adults. The test yields both an abstract score (Ab) and a concrete score (Co). An adjusted raw score is calculated using the formula (Ab - Co + 10) which combines the discriminatory power of both measures and eliminates minus quantities by making zero the lowest possible score. Scores based on a normal population have a mean of 30.7 and a standard deviation of 9.8. The split-half reliability coefficient of the multiple choice form is .88 (Ab score). Alternate form reliabilities range from .71 to .79 (Ab score). Validity studies reported in the clinical manual appear to indicate that the Proverbs Test is a valid instrument for assessing abstract and concrete reasoning.

The Arlin Test of Formal Reasoning consists of 32 items organized into eight subtests which assess an individual's ability to use the eight formal concepts associated with the formal operational stage. The eight formal concepts are: (a) Multiplicative Compensations, (b) Correlations, (c) Probability, (d) Combinational Reasoning, (e) Proportional Reasoning, (f) Forms of Conservation Beyond Direct Verification,

(g) Mechanical Equilibrium, and (h) The Coordination of Two or More Systems or Frames of Reference. The test yields a total raw score and eight subtest raw scores. The total score means range from 13.59 (SD = 4.31) for grades 6-8 to 18.33 (SD = 5.11) for grades 10-12. Test-retest reliabilities ranged from .76 to .89 depending on the version of the test and the time period between testing. Coefficient alphas for the total test composite ranged from .60 to .73. Coefficient alpha is a more general reliability coefficient that can describe the variance of items, whether or not they are in a right-wrong format (Kaplan & Saccuzzo, 1982). Multi-trait, multi-method validity studies reported in the clinical manual and in Arlin (1982) indicate the Arlin Test of Formal Reasoning is a valid test of formal operational thought.

Procedure

The Arlin Test of Formal Reasoning (ATFR) and the Proverbs Test (PT) Best Answer Form were administered to each subject. As both tests are power tests, no time limit was set. The tests were administered by the author of the study and two graduate teaching assistants in the psychology department (instuctors of the classes used for this study). The tests were administered with strict adherence to the respective test manuals.

No effort was made to counterbalance the administration of the tests. The tests were scored by the author of the study with the aid of stencils provided with each test.

The data were analyzed using the Pearson product-moment correlational technique. A Fisher's table was used to establish significance. A series of *t*-tests was used to determine whether significant differences existed between scores on the tests with reference to gender. Means and standard deviations were run on the test score data obtained.

CHAPTER 3

Results

The purpose of this study was to examine the relationships between Arlin Test of Formal Reasoning (ATFR) scores and Proverbs Test (PT) scores of a sample of college students ages 18-20. Proverbs Test scores, Abstract, Concrete, and Adjusted scores, were correlated with the Arlin Test of Formal Reasoning Total score. All Proverbs Test scores were also correlated with the eight subtest scores of the Arlin Test of Formal Reasoning. Those subtests are: Subtest I, Multiplicative Compensations; Subtest II, Correlations; Subtest III, Probability; Subtest IV, Combinational Reasoning; Subtest V, Proportional Reasoning; Subtest VI, Forms of Conservation Beyond Direct Verification; Subtest VII, Mechanical Equilibrium; and Subtest VIII, The Coordination of Two or More Systems or Frames of Reference. The relationships between test scores were determined by using the Pearson product-moment correlational technique.

Also of interest was the determination of significant differences in test scores based on gender. A series of *t*-tests was used to determine the differences between means on the tests.

Finally, the percentages of college students ages 18-20 at each of the five cognitive levels assessed by the Arlin Test of Formal Reasoning was determined. Those levels are: Concrete, High Concrete, Transitional, Low Formal, and High Formal. Further, the determination of the percentages of students operating at formal levels (Low Formal + High Formal) was undertaken.

Descriptive statistics were run on the data collected in this study. The mean scores and standard deviations for the Arlin Test of Formal Reasoning, its eight subtests, and the Proverbs Test are presented in Table 1.

Table 1

Means and Standard Deviations of ATFR and PT Scores

Test		M	SD
ATFR Total		17.39	5.10
I	Multiplicative Compensations	2.46	1.31
II	Correlations	2.28	.95
III	Probability	3.09	1.15
IV	Combinational Reasoning	1.36	1.05
V	Proportional Reasoning	2.08	1.46
VI	Forms of Conservation	1.34	1.31
VII	Mechanical Equalibrium	2.45	1.22
VIII	Frames of Reference	1.75	1.24
PT Abstract		27.59	4.88
PT Concrete		.61	1.09
PT Adjusted		36.80	5.45

Note. ATFR = Arlin Test of Formal Reasoning, PT = Proverbs Test.

The correlational analyses of score data obtained from the Arlin Test of Formal Reasoning Total score and the Proverbs Test Abstract and Adjusted scores indicate that a significant positive correlation exists between the two measures. The correlation coefficient between the Arlin Test of Formal Reasoning and the Proverbs Test Abstract score was .56, as was the correlation coefficient between the Arlin Test of Formal Reasoning Total score and the Proverbs Test Adjusted score ($p < .05$). Tables 2 and 2a provide a summary of the correlations of all test scores.

Table 2

Correlations Between ATFR, ATFR Subtests, and the PT

		ATFR	I	II	III	IV	V
ATFR		1.00					
I	Multiplicative Compensations	.57*	1.00				
II	Correlations	.48*	.21	1.00			
III	Probability	.57*	.18	.11	1.00		
IV	Combinational Reasoning	.39*	.09	.26	.17	1.00	
V	Proportional Reasoning	.69*	.37*	.27*	.08	.18	1.00
VI	Forms of Conservation	.53*	.26	.21	.40*	.20	.19
VII	Mechanical Equilibrium	.38*	-.03	-.01	.30*	-.15	.20
VIII	Frames of Reference	.54*	.22	.14	.19	.43*	-.06
PT	Abstract	.56*	.20	.41*	.25	.33*	.42
PT	Concrete	-.30*	-.14	.00	-.34*	-.02	-.22
PT	Adjusted	.56*	.20	.40*	.35*	.31*	.38*

*$p < .05$

Note. Roman numerals indicate ATFR Subtests.

Table 2a

Correlations Between ATFR, ATFR Subtests, and the PT (Cont'.)

	VI	VII	VIII	PTAB	PTC	PTAD
VI Forms of Conservation	1.00					
VII Mechanical Equilibrium	-.03	1.00				
VIII Frames of Reference	.21	.24	1.00			
PT Abstract	.46*	.06	.21	1.00		
PT Concrete	-.22	-.22	-.06	-.34*	1.00	
PT Adjusted	.49*	.08	.17	.95	-.48	1.00

*$p < .05$.

Note. PTAB = PT Abstract, PTC = PT Concrete, PTAD = PT Adjusted.

A significant negative correlation exists between the Arlin Test of Formal Reasoning Total score (ATFR) and the Proverbs Test Concrete score (PTC), ($r = -.30$, $p < .50$). The Proverbs Test Abstract score (PTAD) correlated significantly with scores on the Arlin Test of Formal Reasoning Subtests II, Correlations; IV,

Combinational Reasoning; V, Proportional Reasoning; and VI, Forms of Conservation Beyond Direct Verification, and were .41, .33, .42, .46, and .40, .31, .38, and .49 respectively. The Proverbs Test Adjusted score also correlated significantly with the score on the Arlin Test of Formal Reasoning Subtest III, Probability (r = .35). The Proverbs Test Concrete score (PTC) was negatively correlated with the score on the Arlin Test of Formal Reasoning Subtest III, Probability (r = -.34). Remaining test and subtest scores were not significantly correlated.

Student's t-tests were also calculated to determine if statistically significant differences occurred between mean scores of male and female subjects on any of the tests. Table 3 contains the results of mean difference tests for the subject pool when broken on the basis of gender.

Table 3

Means, Standard Deviations, and t values of ATFR, ATFR Subtests, and PT Scores for Males and Females

Test		Males (N = 17) M	Males (N = 17) SD	Females (N = 39) M	Females (N = 39) SD	t
ATFR	Total	19.76	5.11	16.36	4.80	2.39
I	Multiplicative Compensations	3.00	1.28	2.23	1.27	2.09
II	Correlations	2.82	1.02	2.90	.94	
III	Probability	3.06	1.30	3.10	1.10	
IV	Combinational Reasoning	1.29	1.31	1.38	.94	
V	Proportional Reasoning	3.00	1.17	1.67	1.40	3.43
VI	Forms of Conservation	1.41	1.50	1.31	1.24	
VII	Mechanical Equilibrium	2.53	1.33	2.41	1.87	
VIII	Frames of Reference	2.65	1.06	1.36	1.11	4.04
PT	Abstract	28.59	3.62	27.15	5.32	
PT	Concrete	.35	.61	.72	1.23	
PT	Adjusted	37.65	4.44	36.44	5.84	

A significant difference exists between male and female scores on the Arlin Test of Formal Reasoning. This difference appears to be due to significant differences between male and female scores on Subtests I, Multiplicative Compensations; V, Proportional Reasoning; and VIII, Frames of Reference.

No significant differences between male and female group means were found on the Proverbs Test. Score means of both groups were nearly equal.

Subjects were classified into five cognitive levels as determined by the Arlin Test of Formal Reasoning. A Summary of the data concerning cognitive levels is presented in Table 4.

Table 4

Number of Students in Cognitive Levels of the ATFR

Level	Male	Female
Concrete	0	1
High Concrete	3	12
Transitional	4	10
Low Formal	5	14
High Formal	5	2

Table 5 provides a summary of the data concerning the percentage of subjects falling into each of the five cognitive levels.

Table 5

Numbers and Percentages of Students at ATFR Cognitive Levels

Level	N	%
Concrete	1	1.8
High Concrete	15	26.8
Transitional	14	25.0
Low Formal	19	33.9
High Formal	7	12.5

N = 56.

The percentage of college students ages 18-20 in this study showing evidence of performing at formal operational levels (Low Formal + High Formal) was 46.4. The percentage of males at these levels was 58.8, while the percentage of females was 41.0.

CHAPTER 4

Discussion

A review of the results of this study indicates that a significant relationship exists between the Arlin Test of Formal Reasoning (ATFR) and the Proverbs Test (PT). However, the Proverbs Test scores correlated significantly with only about half of the eight ATFR subtest scores. The Proverbs Test Abstract score (PTAB) measures an individual's ability to convert the concrete symbols of a proverb into verbal abstractions, in other words, the ability to think abstractly. As this score correlated significantly with Arlin Test of Formal Reasoning Subtests II, Correlations; IV, Combinational Reasoning; and VI, Forms of Conservation Beyond Direct Verification, the implications are that the abilities assessed by these particular subtests may be related to the process of proverb interpretation. Also implied is the possibility that the Arlin Test of Formal Reasoning and the Proverbs Test may be measuring similar, but not necessarily the same abilities. As no scores on the Proverbs Test (PTAB, PTC or PTAD) correlated significantly with scores on Arlin Test of Formal Reasoning Subtests I, Multiplicative Compensations; VII, Mechanical Equilibrium; or VIII, Frames of Reference, it is suggested that the Proverbs

Test and these Arlin Test of Formal Reasoning subtests may not be measuring similar abilities.

The Proverbs Test Adjusted score (PTAD) correlated significantly with the same Arlin Test of Formal Reasoning subtests that correlated significantly with the Proverbs Test Abstract score (PTAB), with the addition of the Arlin Test of Formal Reasoning Subtest III, Probability. The Proverbs Concrete score (PTC) was also negatively correlated with this subtest. As the Proverbs Test Adjusted score contains both the Abstract and Concrete scores (AB - CO + 10), this relationship was to be expected. The Proverbs Test score (PTC) correlated negatively with the Arlin Test of Formal Reasoning Total score (r = -.30) as well as negatively with the Proverbs Abstract and Adjusted scores. This suggests that, as Gorham (1975) stated, they may be measuring opposite abilities--concrete and abstract thinking.

Whereas Billow (1975) was unable to find a significant relationship between proverb interpretation and a combinatorial reasoning task (a formal task), the results of this study show a significant relationship between proverb interpretation as assessed by the Proverbs Test and a written combinational reasoning problem (ATFR Subtest IV, Combinational Reasoning).

Further exploration into the reasons for this difference may need to be undertaken. What portion of the difference lies in the form of the task, whether manipulative or written?

Some light may have been shed on the use of proportional reasoning to interpret proverbs, as this subtest (ATFR Subtest V, Proportional Reasoning) correlated significantly with the Proverbs Test Abstract score. Arlin (1984) states that: "the proportional reasoning concept appears to be a 'necessary but not sufficient condition' for interpreting analogies and expanded metaphors beyond their most literal level" (p. 11). Results of this study appear to support this reasoning.

Findings of this study related to the difference in performance on the Arlin Test of Formal Reasoning and the Proverbs Test by males and females tend to dispute the contention that males and females tend to score relatively the same on written tests of formal thinking. Males in this study scored significantly higher than females on the ATFR. However, evidence does support relatively equal levels of performance by males and females on the Proverbs Test. Evidence from this study also supports the contention that 50% or less of college and university students may not be

performing at formal levels. The percentage of college freshmen in this study performing at formal levels was 46.

The Arlin Test of Formal Reasoning Total score and the Proverbs Test Abstract and Adjusted scores were significantly correlated (r = .56 and r = .56), supporting the use of the Proverbs Test as a measure of formal operations when a quick screening measure may be appropriate. Due to the somewhat lower correlation between the two tests, it is recommended that the Proverbs Test be used as a group test rather than for individual classification. The Proverbs Test could then be used to gain an understanding of the general classroom level of cognitive functioning. Further testing with a measure such as the Arlin Test of Formal Reasoning could then be used for individual evaluation and classification into cognitive levels. Caution should be exercised when generalizing the results of this study, due to the small sample size and the limited age and intellectual range.

References

Ahlawat, K. S. & Billeh, V. Y. (1982). The factor structure of the Longeot Test: a measure of logical thinking. Journal of Research in Science Teaching, 19(8), 647-658.

Arlin, P. K. (1982). A multitrait-multimethod validity study of a test of formal reasoning. Educational and Psychological Measurement, 42, 1077-1088.

Arlin, P. K. (1984a). Arlin Test of Formal Reasoning. East Aurora, New York: Slosson Educational Publications.

Arlin, P. K. (1984b). Arlin Test of Formal Reasoning: Test Manual. East Aurora, New York: Slosson Educational Publications.

Barnes, G. (1977). Scores on a Piaget-type questionnaire versus semester grades for lower-division college physics students. American Journal of Physics, 45(9), 841-847.

Bart, W. M. (1971). The factor structure of formal operations. British Journal of Educational Psychology, 41, 70-77.

Bender, D. S. & Milakofsky, L. (1982). College chemistry and Piaget: the relationship of aptitude and achievement measures. Journal of Research in Science Teaching, 19(3), 205-216.

Billow, R. M. (1975). A cognitive developmental study of metaphor comprehension. Developmental Psychology, 11(4), 415-423.

Blake, A. J. D.; Lawson, A. E. & Nordland F. H. (1976). The Karplus Islands Puzzle: does it measure Piagetian operations? Journal of Research in Science Teaching, 13(5), 397-404.

Danner, F. W. & Day, C. M. (1977). Eliciting formal operations. Child Development, 48(4), 1600-1606.

Elmore, C. M. & Gorham, D. R. (1957). Measuring the impairment of the abstracting function with the Proverbs Test. Journal of Clinical Psychology, 13 (3), 263-266.

Fogel, M. L. (1965). The Proverbs Test in the appraisal of cerebral disease. Journal of Genetic Psychology, 72, 269-275.

Fogelman, K. R. (1972). Piagetian tests and sex differences II. Educational Research, 12(2), 144-155.

Gorham, D. R. (1956a). Use of the Proverbs Test for differentiating schizophrenics from normals. Journal of Consulting Psychology, 20(6), 435-440.

Gorham, D. R. (1956b). A Proverbs Test for clinical and experimental use. Psychological Reports, Monograph Supplement 1, 2, 1-12.

Gorham, D. R. (1956c). The Proverbs Test. Missoula, MT: Psychological Test Specialists.

Gorham, D. R. (1975). Additional norms and scoring suggestions for the Proverbs Test. Psychological Reports, 13, 487-492.

Gorham, D. R. (1975). Clinical Manual for the Proverbs Test. Missoula, MT: Psychological Test Specialists.

Hardy-Brown, K. (1979). Formal operations and the issue of generalizability: the analysis of poetry by college students. Human Development, 22, 127-136.

Inhelder, B & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. USA: Basic Books, Inc.

Jurjevick, R. M. (1967). Intellectual assessment with Gorham's Proverbs Test, Raven's Progressive Matrices and WAIS. Psychological Reports, 20, 1285-1286.

Karplus, E. F. & Karplus, R. (1970). Intellectual development beyond elementary school. School Science and Mathematics, 70, 398-406.

Lawson, A. E. (1985). A review of research on formal reasoning and science teaching. Journal of Research in Science Teaching, 22(7), 569-617.

Kaplan, R. M. & Saccuzzo, D. P. (1982). Psychological testing: principles and issues. Monterey: CA.

Lawson, A. E. (1979). Relationships among performance on group administered items of formal reasoning. Perceptual and Motor Skills, 48, 71-78.

Lawson, A. E. (1978). The development and validation of a classroom test of formal Reasoning. Journal of Research in Science Teaching, 15(11), 11-24.

Lawson, A. E. (1975). Sex differences in concrete and formal reasoning ability as measured by manipulative tasks and written tests. Science Education, 59(3), 397-405.

Lawson, A. E. & Blake, A. J. D. (1976). Concrete and formal thinking abilities in high school biology students as measured by three separate instruments. Journal of Research in Science Teaching, 13(3), 227-235.

Lawson, A. E.; Nordland, F. H.; & DeVito, A. (1974). Piagetian tasks: a crossover study of learning effect and reliability. Science Education, 58, 267-276.

Lasky, L. (1967). Alternate forms of the multiple choice Proverbs test. The Journal of Psychology, 65, 59-60.

Lovell, K. (1971). Some problems associated with formal thought and its measurement. In D. R. Green, M. P. Ford & G. B. Flamer (Eds.). Measurement and Piaget. (pp. 81-93). New York: McGraw-Hill Book Co.

Martin, W. T. (1967). Analysis of the abstracting function in reasoning using an experimental test. Psychological Reports, 21, 593-598.

Matarazzo, J. D. (1972). Wechsler's measurement and appraisal of adult intelligence (5th ed.). New York: Oxford University Press.

McKinnon, J. W. & Renner, J.W. (1971). Are colleges concerned with intellectual development? American Journal of Physics, 39, 1047-1052.

Patterson, H. O. & Milakofsky, L. (1980). A paper-and-pencil inventory for the assessment of Piaget's tasks. Applied Psychological Measurement, 4, 341-353.

Piaget, J. (1923). La pensée symbolique et la pensée de l'enfant. Archives de Psychologie, 18, 273-304.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. Human Development, 15, 1.12.

Raven, R. J. (1973). The development of a test of Piaget's logical operations. Science Education, 57(3), 377-385.

Renner, J.W. & Paske, W. C. (1977). Comparing two forms of instruction in college physics. Americal Journal of Physics, 45(9), 851-859.

Richardson, C. & Church, J. (1959). A developmental analysis of proverb interpretation. The Journal Genetic Psychology, 94, 169-179.

Ross, R. J. (1974). The empirical status of the formal operations. Adolescence, 9, 413-420.

Rowell, J. A. & Hoffman, J. P. (1975). Group tests for distinguishing formal from concrete thinkers. Journal of Research in Science Teaching, 12(2), 157-164.

Sayre, S. & Ball, D. W. (1975). Piagetian cognitive development and achievement in science. Journal of Research in Science Teaching, 12(2), 165-174.

Schneider, L. S. & Renner, J. W. (1980). Concrete and formal teaching. Journal of Research in Science Teaching, 17(6), 503-517.

Schwebel, M. (1975). Formal operations in first-year college students. The Journal of Psychology, 91, 133-141.

Shayer, M.; Adey, P. & Wylam, H. (1981). Group tests of cognitive development: ideals and realization. Journal of Research in Science Teaching, 18(2), 157-168.

Shayer, M. & Wharry, D. (1973). Piaget in the classroom I: testing of a whole class at the same time. School Science, 54(191), 447-457.

Smith, R. C. (1971). The use of the Proverbs Test for the identification of psychotic disorder. Journal of Clinical Psychology, 27(2), 277.

Staver, J. R. & Gabel, D. L. (1979). The development and construct validation of a group administered test of formal thought. Journal of Research in Science Teaching, 16(6). 535-544.

Tisher, R. P. (1971). A Piagetian questionnaire applied to pupils in a secondary school. Child Development, 42, 1633-1636.

Tobin, K. G. & Capie, W. (1981). The development of a group test of logical thinking. Educational and Psychological Measurement, 41, 413-423.

Tomlinson-Keasey, C. (1972). Formal operations in females from eleven to fifty-four years of age. Developmental Psychology, 6(2). 364.

Tschopp, J. K. & Kurdek, L. A. (1981). An Assessment of the relation between traditional and pencil-and-paper formal operations tasks. Journal of Research in Science Teaching, 18(1), 87-91.

Walker, R. A.; Hendrix, J. R. & Mertens, T. R. (1979). Written Piagetian Task Instrument: its development and use. Science Education, 63(2), 211-220.

APPENDIX A

Consent Form

I, _____,do agree to participate in a study conducted by Cindy Katherine Primeau as partial fulfillment of her graduate program at Emporia State University. I understand that I will be asked to take two tests, the Arlin Test of Formal Reasoning and the Gorham's Proverb Test, and that the scores on these tests will be used in the study. I have also been informed that the results of my tests will be kept strictly confidential, and that I have the right to obtain the results of my tests at the completion of the study. I also understand that I may withdraw from this study at any time if I should, for any reason, decide that I do not want the results of my tests included.

_____

Signature of the Testee

_____

Date