AN ABSTRACT OF THE THESIS OF

Glenda Jean Young for the Master of Science

in Clinical Psychology presented on October, 1993

Title: Relationships Between the Arlin Test of Formal Reasoning and the Wechsler Adult Intelligence Scale-Revised

Abstract approved:

The purpose of this study was to investigate the relationships between the Arlin Test of Formal Reasoning (ATFR) and the Wechsler Adult Intelligence Scale-Revised (WAIS-R). Additional issues explored were the percentage of the sample group operating at formal operational levels and whether there were significant differences between scores on the two tests that were related to differences in gender. The sample consisted of 20 college students (9 females and 11 males) ranging in age from 18 to 20. Both the ATFR and the WAIS-R were administered to each subject. Scores for both tests were recorded for each subject as were dates of birth and gender.

Pearson product-moment correlation coefficients were calculated to determine the relationships between the ATFR Total scores and the WAIS-R Full Scale, Verbal, and Performance IQ scores and the 11 WAIS-R subtest scores. A series of *t*-tests was computed to determine significant differences in performance on the two tests by males and females. Subjects were categorized by the five cognitive levels assessed by the ATFR and the percentage of students

operating at each of the five levels was calculated.

The ATFR total scores correlated significantly with the WAIS-R Full Scale IQ score, Verbal IQ score, and Information and Vocabulary subtest scores. No significant differences between male and female group means were found on the ATFR or WAIS-R scores. Seventy percent of the subjects in this sample group were found to be operating at a formal level of cognitive ability, with 15% of the sample operating at a transitional level between formal and concrete cognitive ability and the remaining 15% assessed at a concrete level of cognition.

The statistically significant relationships between the Information and Vocabulary subtests scores of the WAIS-R and the ATFR in this sample group suggest they have a greater potential for measuring formal reasoning ability than do other component parts of the WAIS-R. Further research is indicated to define the nature of this relationship. Limitations in the generalizability of this study, due to the small sample size and timing of data collection, also warrant additional research to substantiate the findings of this study.

Relationships Between The Arlin Test Of Formal Reasoning And The Wechsler Adult Intelligence Scale-Revised

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

Glenda Young

October, 1993

John O. Schwenn

Approved for the Major Division

Faye N. Vowell

Approved for the Graduate Council

ACKNOWLEDGEMENTS

I first wish to express sincere gratitude to my committee members, Chairperson Dr. David Dungan, Mr. Howard Carvajal and Dr. Edward Butler. The expertise, information, support, and encouragement provided by each one will always be greatly appreciated.

Deepest thanks also go to my family, from whom I have received unconditional encouragement and moral support. Finally, I wish to acknowledge and thank my friend Sue Six, with whom I have commiserated the process and celebrated the progress during the course of our respective efforts in thesis writing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

CHAPTER 1

Introduction

Development of intelligence and cognitive theories have been major considerations for psychological research for decades. Efforts to define intelligence have been extensive but have not lead to a universally accepted definition of intelligence. Therefore, intelligence remains one of the most controversial concepts in psychology (Weinberg, 1989). The inability to satisfactorily define intelligence has not inhibited the research leading to a number of theories. Among the most prevalent of these are two theories of intelligence relevant to this study, psychometric and cognitive developmental.

The concept of intelligence as something that can be quantified emerged in the late 1800's as an attempt to explain differences in the abilities of individuals (Howe, 1989). Work in this area led to the development of instruments to measure or quantify intelligence. This is the psychometric approach to intelligence.

A different perspective of intelligence is closely identified with the work of Jean Piaget. Piaget began to address the question of how knowledge is acquired while working in the laboratories of Alfred Binet, the man credited with originating intelligence testing (Evans, 1973). Piaget's subsequent empirical investigation of this and related questions led to the formulation of a cognitive developmental theory in which he identified four progressively complex age-related cognitive stages of intellectual development: the sensory-motor, preoperational, concrete operational, and the formal operational stages. (Inhelder & Piaget, 1958; Evans, 1973; Piaget, 1972; Brainerd, 1978).

The differences between psychometric and cognitive developmental concepts of intelligence are especially obvious in the methods of intelligence assessment employed by each approach. IQ tests, perhaps the hallmark of psychometric assessment instruments, are based on assumptions of intelligence which compare an individual's response to others of the same age group, while cognitive developmental assessment has traditionally relied on interview and task performance to determine an individual's placement in a sequence of developmental stages (DeVries, 1974). Past attempts to explore relationships between the two approaches have been hampered by the dissimilarity of their assessment methods and have led to debate as to the practical value of such comparisons to research, clinical, and educational needs. Elkind (1971) asserted that disagreement between the two approaches is not because of fundamental differences in belief about the nature of the intelligence, but because each perspective attempts to investigate different facets of intelligent behavior.

Researchers and educators have been working for some time to develop instruments that effectively assess individual development from concrete to formal reasoning (stages three and four in Piagetian theory) in an effort to match developmental levels and educational methods (Nagy & Griffiths, 1982). The resulting instruments, with improved ability to assess individual differences with respect to Piagetian developmental stages, greatly facilitate renewed efforts to further explore relationships of the facets of intelligence represented by psychometric and cognitive developmental intelligence theory. The focus of this research project is a comparison of the relationship of the Piagetian based concept of formal reasoning and the psychometric concept of intelligence (IQ). This will be accomplished by comparing scores from a sample group using the Arlin Test of Formal Reasoning (ATFR) as a measure of cognitive developmental formal reasoning ability and the Wechsler Adult Intelligence Scale-Revised (WAIS-R) as a measure of psychometric IQ.

Literature Review

As noted earlier, efforts to compare formal reasoning ability and IQ have been hindered by the dissimilarity in assessment methods. Exemplary of some of the problems found in early research in this areas is an early study by Furst (1950) who employed instruments developed by faculty members at the University of Chicago to measure intelligence, critical thinking ability, and scholastic achievement in high school students. References relating to the reliability or validity and the objectives of these instruments were addressed only in very general terms. Yet according to the author, the results and conclusions of the study were in large part based on the assumption the tests used in this study were "reliable and valid measures of their particular objectives" (p. 621). Although the deficiencies in this study are blatant and not representative of the quality of most research today, it does serve to point to the importance of establishing the qualification of instruments for the stated purpose of any research project and to identify the objectives of those instruments.

Before addressing these issues as they relate specifically to the two instruments chosen for this study, the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and the Arlin Test of Formal Reasoning (ATFR), a brief overview of the history of intelligence theory is provided. This information as it relates to both psychometric and Piagetian perspectives is relevant to understanding the concepts and influences involved in the development and use of these instruments.

Development of the Psychometric Approach

Early efforts in the development of intelligence theory were concentrated on the laboratory experiments which were directed toward sensory acuity, methods sometimes referred to as "brass instrument psychology" (Thorndike, 1990). The founding of this quantitative approach is most often credited to Gustav Fechner whose book Elemente der Psychophysic published in 1860 applied observations in physics, astronomy, biology, mathematics and physiology to the idea of psychological measurement. Hemholtz followed with empirical emphasis on developing methods of measuring sensory fields of vision and hearing as well as working on reaction-time experiments. However, it was Wilhelm Wundt, known for his establishment of the first psychological laboratory and his extensive work with sensori-motor measurements, who is most often associated with the early development of quantitative psychology (Hunt, 1936).

Other important contributors to the field include Sir Francis Galton whose major contribution was his adaptation and innovation in statistical methods (Guilford, 1936). It was Galton's adaptation of the normal curve and his concept of standard deviation along with his development of correlational statistical methods that provided the means for psychologists to begin to measure aspects of socially related behavior (Evans & Waites, 1981).

Galton's work came to the attention of James McKeen Cattell, who as a young graduate student working with Wilhelm Wundt, became interested in exploring individual differences rather than general features of the mind. Cattell's subsequent collaboration with Galton resulted in research directed to the development of an effective method to measure individual intelligence. Coining the term "mental tests" Cattell, while unsuccessful in developing an instrument that was deemed capable of measuring intelligence, was effective in generating considerable interest in the potentials of such instruments to the psychological community of the time (Fancher, 1985).

The first person credited with a viable intelligence test was Alfred Binet. In 1904 he began developing a method of rating children in an effort to distinguish between children who were underachievers and those with diminished intellectual capacity. Binet's work evolved into an intelligence scale based on a series of graded intellectual tasks that could be effectively measured by evaluating the results in terms of age units, thus introducing the "mental age method" (Wechsler, 1944). Binet based his work on the assumption intelligence was a combination of diverse functions whose arrangements were unique to each individual. By the time of Binet's death in 1911 he, along with Theodore Simon, had revised and improved the intelligence scale considerably, and the Binet-Simon Scale quickly gained the attention of the psychological circles in the western world (Fancher, 1985).

A contemporary of Binet's, Charles Spearman, developed a two factor theory of intelligence composed of a general factor referred to as "*g*" and specific factors or "*s*."

This theory differed from Binet's assumption of intelligence as a collection of diverse functions, as Spearman postulated that "*g*" was an underlying factor in all of Binet's items with "*s*" factors of varying degrees present in each. Spearman's concept provided further rationale for the interpretation of Binet's scale as a measure of intelligence. Although still unproven, this concept remains influential in current psychometric theory (Fancher, 1985). Wechsler's (1958) statement "I remain a reformed but unchastened Spearmanite" (p. viii), attests to the impact of Spearman's ideas on the subsequent development of psychometric instruments.

William Stern and Henry Goddard, contemporaries of Binet, were instrumental in the growing interest and development of intelligence tests as influential men in the psychological profession in America, they fostered widespread use of such instruments. William Stern developed the concept of the intelligence quotient or IQ by identifying a ratio between chronological age and mental age. Stern's formula: intelligence quotient (IQ) = mental age divided by chronological age, introduced in 1912, allowed for the quantifying of an individual's difference in performance on standardized tests as a relative rather than absolute difference, thus providing a means of quantifying differences on a one dimensional scale (Stoddard, 1944).

Henry Goddard was a significant but somewhat more controversial influence in the spread of the Binet methods in America. Goddard's book The Kallikak Family, published in 1914, was written to support his theory that a recessive gene was responsible for inferior intelligence. Following this reasoning, Goddard looked to intelligence tests as a means to identify the feebleminded as a first step in preventing the spread of hereditary feeblemindedness. This book, long since discredited, nevertheless demonstrated the ability to relate psychometric instruments to social issues, setting an important precedent for the application of psychometric instruments to numerous social purposes (Fancher, 1985).

Lewis Terman became involved in translating and adapting the Binet-Simon test as well as refining the method of computing IQ scores. His work resulted in the 1916 publication of the Stanford Revision of the Binet-Simon Scale (Terman, 1916), a forerunner of the Stanford-Binet scales in use today.

The advent of World War I gave new impetus to psychological testing and was dictated by the needs of the military to screen incoming troops for placement. A committee headed by Robert Yerkes, drawing largely on Spearman's "*g*" theory and the work of Alfred Binet, rapidly developed tests to meet the needs of the time (Von Mayhauser, 1989). Postwar application of tests developed during the war was extensive and enthusiastic and led to wide spread acceptance of intelligence tests in educational, clinical, and industrial settings. The popularity of testing led to some excesses in expectation and uses of psychometric testing, which reinforced the need to focus on test reliability and validity in subsequent research and test development (Hunt, 1936).

The information concerning the development of psychometric intelligence theory presented here, although far from exhaustive, requires the mention of one more individual who is responsible for intelligence scales widely used and recognized in America today. David Wechsler, a Master's level student of psychology at the beginning of World War I, served as a volunteer test scorer and later as an administrator of individual IQ tests in the army. This was followed by a brief stint in England studying with Spearman (Fancher, 1985). Wechsler's endorsement of Spearman's two factor theory, his pragmatic view of the use of intelligence tests, and his recognition of the need for an individually administered intelligence test for adults, were significant factors in the development of the Wechsler-Bellevue Intelligence Scale (Wechsler, 1944).

Various versions and revisions of the Wechsler-Bellevue are widely used today, a tribute to "Wechsler's original insight and sensitivity" (Zachary, 1990, p. 276). Wechsler's understanding of the practical application of IQ tests is exemplified by his statement that intelligence tests measure "the capacity of an individual to understand the world about him and his resourcefulness to cope with its challenges" (Wechsler, 1975, p. 139).

The Wechsler Adult Intelligence Scale-Revised

The most current version of the adult Wechsler Scales is the Wechsler Adult Intelligence Scale-Revised (WAIS-R) which was published in 1981. The WAIS-R's development can be traced directly from its predecessors, the Wechsler Adult Intelligence Scale (WAIS) published in 1955 and the Wechsler-Bellevue Intelligence Scale, first published in 1939. The content of the WAIS-R retains much of the content and features of the WAIS, with change directed primarily toward content updating, modification of items to reflect changes in item difficulty, and scoring modifications to reflect advances in data analysis (Wechsler, 1981).

Standardization procedures for the WAIS-R were also changed to be more representative of the population it was designed to serve. The norming of this test involved testing equal numbers of men and women between the ages of 16 and 74 over a period of four years following a stratification plan that included the variables of age, sex, race, geographic region, occupation, education, and urban-rural residence. This plan was developed using data from the 1970 United States Census, as well as more recent census data as they became available (Wechsler, 1981).

The subject of test reliability and validity was addressed by Wechsler in the WAIS-R Manual in some detail. For reliability, a split-half procedure was used and the Spearman-Brown formula applied for all but two of the eleven subtests, Digit Span and Digit Symbol, in which a re-test procedure was used. Standard errors of measurement as well as additional reliability information were provided separately by age group with nine divisions delineated between the ages of 16 and 74. Reliability was high across all age groups with an average coefficient of .97 for Full Scale IQs; the overall reliability coefficients of the WAIS-R were consistently high. The validity of the instrument was primarily based on validity studies of its predecessors, the Wechsler-Bellevue and the WAIS, that included comparisons with other established IQ tests, empirical studies of groups of known intellectual level, and factor analytic research (Wechsler, 1981).

The WAIS-R manual was criticized in the Ninth Mental Measurement Year Book (Watkins, 1985) for a lack of validity data, but reports the split-half reliability coefficients as "quite impressive" (p. 1702). Anastasi (1988) comments on the assumption that the WAIS-R can successfully draw on the earlier research of the Wechsler adult scales with regard to validity as the changes introduced by later revisions are improvements on their predecessors and are more likely to underestimate its validity rather than overstate it. Perhaps the statement by Wechsler (1981) that "a body of evidence both rational and empirical attests to the validity of the Wechsler adult scale as a measure of global intelligence" (p. 49) best attests to the reason that the WAIS-R is the individual intelligence test most widely used today (Kline, 1991).

The overall attitudes about the validity and practical use of intelligence tests are positive, but there have been persistent criticisms of intelligence testing almost from their onset (Snyderman & Rothman, 1987). The detractors of IQ testing run the gamut from extreme positions like that of Howe (1989), who describes intelligence as nothing more than a descriptive term and a product of the imagination of twentieth century psychologists, to more specific criticisms such as concerns about the limitation of what the tests actually measure, socioeconomic biases of such tests, and stigmatizing effects for those with low scores (Snyderman & Rothman, 1987). Weinberg (1989) identifies the root of the IQ test controversy in the limitations of what the tests measure, due to the limited ability of IQ tests to provide a representative sample of the full repertoire of human adaptive behavior. More simply stated, there may be more to intelligence than these tests measure.

Cognitive Developmental Approach

Controversy often brings with it suggestions for new directions in research and frequently leads to the development of additional theoretical perspectives. The work of Piaget dominates a theoretical position that is more concerned with the qualitative approach of how we think in contrast to the quantifying approach of the psychometric position with emphasis on attempting to measure what we know (DeVries, 1974).

Piaget began to distinguish himself by publishing his first scientific paper at age 10, and by age 18 he had received his bachelor's degree in biology (Brainerd, 1978). After receiving his Ph.D in biology in 1918, Piaget's interests in philosophy and psychology began to dominate his career and in 1919 he began work with Simon in Binet's laboratory in Paris on standardization of intelligence scales with children. This work led Piaget to explore the reasoning process underlying the children's responses and marked the beginning of his research and formation of his cognitive theory of development (Evans, 1973).

Piaget proposed four age-related stages of development: the sensory-motor (birth to 2 years), the preoperational period (2 to 7 years), concrete operational (7 to 11 years), and the formal operational period (11 to 15 years). The cognitive development through these stages is directional, always progressing from the simple to the complex (Piaget & Inhelder, 1969). Formal reasoning has been defined as the ability to engage in abstract thought utilizing a hypothetic-deductive process. This process facilitates problem-solving through the generation of hypotheses which are validated or rejected through logical analysis and empirical investigation (Arlin, 1975). The assessment of formal reasoning has became the focus of extensive research particularly with respect to educational applicability. Educators, influenced by Piaget's theory, began to explore a relationship between an individual's level of cognitive development and academic achievement, hypothesizing that matching instructional methods to developmental levels could potentially improve academic performance (Lawson, 1978; Nagy & Griffiths, 1982; Strahan & O'Sullivan, 1988). However, the clinical method employed by Piaget to assess cognitive development, particularly of formal reasoning capabilities, was time-consuming, requiring the use of cumbersome laboratory equipment and trained evaluators (Inhelder & Piaget, 1958). Although subsequent studies tended to validate the use of Piagetian methods as reliable measures of formal reasoning (Elkind, 1961; Bart, 1971; DeVries, 1974), their use as a tool for extensive research was limited. This recognition gave impetus to the attempt to develop more efficient assessment methods based on Piagetian tasks in the form of paper and pencil group tests (Staver & Gabel, 1979; Shayer, Adey, & Wylam, 1981; Arlin, 1982).

Tests of Cognitive Development

One effort to develop a group pencil and paper test was undertaken by Raven (1973), who designed the Raven's Test of Logical Operations (RTLO) using Piagetian problem solving rules but not Piaget's tasks. Using a multiple choice format, problems were presented with pictorial representation followed by printed questions; solution choices were also in pictorial form. The RTLO was administered to 424 students in grades 3, 5, 7, and 9. Raven's analysis of this study concluded this instrument had met his criteria as a reliable and practical instrument for use in determining "the developmental patterns of logical operations in children across grade levels" (p. 384). A later review of this study questioned the validity evidence stating that no attempt had been made to compare Raven's instrument with traditional tasks as well as pointing out deficiencies in the statistical analysis in the study (Nagy & Griffiths, 1982).

Rowell and Hoffman (1975) criticized paper and pencil methods to measure developmental level as not being sufficiently true to Piagetian tasks and focused their efforts to develop a group test using Piagetian type tasks modified to afford easy administration and scoring. Mediated by the need for tasks of different subject matter but of like structure that would adapt to classroom requirements, they chose Inhelder and Piaget's (1958) chemical change experiment and the pendulum experiment. The sample of students, ages 12 to 16 from a South Australian metropolitan high school, was provided with written instructions, worksheets, and the needed apparatus and were prepared by reading through their instructions with an instructor. While the results of this study indicated a significantly high correlation coefficient (r = .56) to suggest the two tasks provide the same measure of developmental level and supported adaption of Piagetian tasks to group settings, the difficulties inherent in standardizing this type of assessment for extensive research were prohibitive.

The development of the Karplus Islands Puzzle is also representative of the efforts to develop an effective tool for assessment of formal reasoning ability. The Islands' puzzle consisted of a printed map of four islands identified by a letter or name and a set of three questions designed to elicit deductive reasoning indicative of formal operational ability. Introduction of the puzzle and clues were presented orally by an instructor with visual reference made to a blackboard representation of the map. Subjects were instructed to write answers and explanations of their answers on the test sheet. The subjects were then placed in one of six developmental categories based on the quality of their responses (Karplus & Karplus 1970).

A further study of this instrument was done by Blake, Lawson and Nordland (1976) in which the performance of 126 high school students on Karplus Islands Puzzle was compared to their performance on three Piagetian tasks: conservation of volume, bending rods, and balance beam (Inhelder &

Piaget, 1958). The results pointed to limitations in the instrument itself and found only a moderate relationship between the Islands Puzzle and the Piagetian tasks. They concluded the Karplus Islands Puzzle's reliance on one task limited its effectiveness in characterizing developmental levels. Together with the problems involved in standardized training of instructors to administer and categorize the test this instrument fell short of research needs.

Lawson (1978), not convinced of the value of strictly pencil and paper measures, used a method in which materials were used by the investigator to demonstrate situations to a group of individuals who then responded in writing to questions posed by the investigator. Lawson argued that the method used in his Classroom Test of Formal Reasoning (CTFR) kept the motivational aspects of the clinical interview with the added practical aspects of group administration.

The CTFR consisted of 15 items representing a cross section of formal operational skills. Each item required a demonstration involving the use of physical materials or apparatus. The questions posed by the demonstrations were also presented in written form in individual test booklets along with a list of possible answers. Subjects were required to provide an explanation for their choice of answers in written form in the test booklet. The 513 students selected for this study were from a population of grades 8-12 in two suburban communities in the San Francisco Bay area. Lawson's conclusion from this study was optimistic. He reported the test to compare favorably to classical Piagetian tasks as a measure of formal reasoning. He further stated that the CTFR could be administered to groups in a short period of time with minimal scoring time needed.

Succeeding studies of the CTFR have not entirely substantiated Lawson's conclusions. Pratt and Hacker (1984) questioned the validity of both the CTFR and a modified version of the CTFR. They cited problems in the test format that failed to reflect some essential aspects of formal reasoning that were identified in the work of Inhelder and Piaget (1958), such as the ability to contemplate a range of hypotheses for any given problem. A more recent study by Hacker (1989) further qualifies the CTFR deficiencies questioning the validity of the instrument as a measure of the unitary formal reasoning construct.

The Test of Logical Thinking (TOLT) was developed by Tobin and Capie (1981) from ten of the items used in the aforementioned study by Lawson (1978). This test incorporated the use of a color-video tape to demonstrate the situations in an effort to standardize administration procedures. In its final form, the TOLT provided multiple choices for a correct response as well as multiple choice justification statements for each item. This study was optimistically regarded by its authors as having high test reliability which was indicated by a high level of internal consistency with supportive evidence of criterion-related validity. However, the criticisms of the CTFR also apply to the TOLT as they were both developed from Lawson's previous research. A more recent study comparing the psychometric properties of the TOLT, the CTFR, and the Longeot Test of Logical Thinking, gave additional indications that all three of these instruments are lacking in concurrent validity (Ahlawat & Billeh, 1987).

A number of more comprehensive instruments have subsequently been developed through the conversion of Piagetian type tasks to written form. Included among these are the Piagetian Task Instrument (PTI) described as a set of problems requiring recognizable reasoning patterns to achieve solution (Walker, Hendrix, & Mertens, 1979). Another such instrument is the Piagetian Logical Operations Test (PLOT), which is described as an objective multiple choice test consisting of four scales that correspond to Piagetian traits of formal thought (Staver & Gabel, 1979).

The written PTI consisted of six tasks, two items for each of the following three reasoning patterns: propositional logic, combinatorial logic, and hypothetico-deductive logic. These tasks were based on the original work of Inhelder and Piaget (1958) but substituted the science-related content for more familiar components to lessen intimidation and anxiety for the test takers. The test was administered to 86 genetics students between the ages of 18-27. Seventy-two percent of this sample were classified as operating at the formal level, and 28% were classified as nonformal by the PTI. The authors of this study, Walker, Hendrix, and Mertens (1979) concluded that their results supported the use of this instrument to assess formal reasoning levels for the three reasoning patterns considered.

Although far from being an exhaustive list of the group paper and pencil tests designed to assess attainment of formal reasoning based on Piagetian tasks and clinical interview techniques, the above mentioned studies are representative of the scope and direction of this area of research. Nagy and Griffiths (1982) reviewed a number of studies of paper and pencil group tests including the PTI, whose main flaw, in their view, was a failure of the instrument to detect any relationship between age and developmental level. The body of research review by Nagy and Griffiths includes five of the seven instruments mentioned above as well as numerous other methods and instruments based on Piaget's theory. The authors concluded that "effective group tests have yet to be developed" (p. 548). Despite this conclusion, the point was also made that the continued development of group tests of this nature is an important avenue of research especially when the need for large numbers of subjects are called for as is certainly the case for research of aptitude-treatment interactions.

Efforts to increase the reliability and validity of paper and pencil instruments for the measurement of formal reasoning ability has continued, and research by Roberge and Flexer (1982) suggests the test they developed, the Formal Operational Reasoning Test (FORT) is both reliable and valid. However, the FORT falls short in the area of comprehensiveness of assessment as it measures only three formal operations skills. This criticism is supported by the Nagy and Griffiths (1982) study which criticizes several instruments similar to the FORT for limitations due to the use of fewer than the eight formal schemes that make up the formal reasoning construct.

The Arlin Test of Formal Reasoning

A test instrument identified in a review by Fakouri (1985) as "a great improvement over its predecessors and a welcome addition to assessment instruments for professionals who are engaged in educational and psychological assessment and research" (p. 43) is the Arlin Test of Formal Reasoning (ATFR). Items used in the ATFR were selected to parallel closely the eight concepts of formal reasoning initially employed by Inhelder and Piaget (1958). However, the ATFR items differ from Piagetian prototypes in that they are represented in non-scientific and non-mathematical terminology. Efforts were also made to produce items that were independent of middle school and high school course curricula, and illustrations were chosen to be representative of items familiar to North American school children.

Reliability and validity data on the ATFR are presented in a multitrait-multimethod study (Arlin, 1982) using a random sample of 38 military recruits from a population of 394 recruits who were tested at an east coast training center. An earlier version of the ATFR and Piagetian clinical interview representing the same formal construct schemas was used. Although the mutitrait-multimethod matrix did not show all the validity and reliability coefficients to be above the desired .80, the analysis did show consistently high relationships between six traits measured by both methods, ranging from .55 to .74, which indicated convergent validity. Discriminant validity was demonstrated by the comparison of correlation coefficients represented in the matrix and were concluded to be indicative of "a highly significant general level of validity" (p. 1086). A review of this validity study in the ATFR manual (Arlin, 1984) reported test-retest reliability as ranging from .76 to .89 with differences ascribed to the use of different versions of the ATFR and time periods between testings.

In addition to the positive review by Fakouri (1985) mentioned earlier, Santmire's (1985) review of the ATFR in the <u>Ninth Mental Measurements Yearbook</u> is also generally optimistic about the instrument. Some concern was expressed in this review for the consistency with which the ATFR corresponds to the levels of thinking developed by Inhelder and Piaget (1958), especially as it relates to concrete thinking, the lowest level of the ATFR. However, the total score assessment holds up well as a measure of formal reasoning and was characterized as "a step in the right direction" (p. 83).

The ATFR is also practical for research use as it provides for ease of administration that requires no special training, has the capacity to test large groups in 30 to 45 minutes, and has an objective uncomplicated scoring system. For these reasons as well as the ATFR's comparatively high degree of reliability and validity as a tool to assess formal operational ability, the ATFR was chosen for this study.

Purpose of this Study

Research directed towards comparison of psychometric and cognitive developmental intelligence have resulted in mixed results showing low and moderate relationships between the two with indications by factor analytic studies supporting the idea that the two types of instruments are measures of separate facets of intelligence (DeVries, 1974). Therefore, this study does not attempt to determine if the WAIS-R and the ATFR both measure intelligence, but rather its purpose is to examine the relationship between formal reasoning ability, a measure of cognitive developmental level, and IQ, a measure of general intellectual ability. This study focuses on a comparison of the scores of the ATFR to the scores achieved by the same subjects on the Verbal IQ, Performance IQ, and Full Scale IQ scores, as well as with the scores of each subtest of the WAIS-R.

It is hypothesized that those individuals who receive high scores on the ATFR indicative of formal reasoning ability, according to the assignment of raw ATFR scores to cognitive levels in the ATFR manual (Arlin, 1984), will achieve higher Full Scale IQ scores than those who are designated as operating at the concrete level by the ATFR. It is further suggested there may be more subtle differences in IQ performance between those operating at the formal and concrete levels that are not discernable in the comparison of the ATFR scores to WAIS-R Full Scale Scores which could become apparent by comparing the ATFR scores with the Performance and Verbal scales and the subtest scores of the WAIS-R.

Significance of this Study

Group administered paper and pencil tests have been much in evidence in recent research that refutes Piaget's assertion that formal reasoning ability generally becomes apparent between the ages of 11 to 15 (Piaget & Inhelder, 1969). Instead it has been found that the age formal reasoning ability attained varies greatly, with a large percentage of teenagers and young adults found to be still operating in the concrete operational stage of development (McKinnon & Renner, 1971; Lawson & Blake, 1976; Chiappetta, 1976). Studies targeting college students who are identified as operating at a concrete level have had some success demonstrating that matching instructional methods, with emphasis on application of formal reasoning strategies to developmental level, could improve scholastic achievement (Danner & Day, 1977; Bender & Milakofsky, 1982; Niaz, 1987).

The percentage of college students in the sample group chosen for this study found by the results of the ATFR to be operating at a concrete level will either lend support to or detract from the research which indicates that up to 52% of the traditional age college population remain at the concrete operational level (Chiappetta, 1976). In addition, results of this study, which point to differences in IQ performance by college students that is related to their level of cognitive development, could provide information as to how formal reasoning ability relates to increased intellectual performance. This could lead to suggestions for the development of educational tools designed to foster optimum academic achievement by matching educational methods to individual needs. As a final note, evidence of more specific differences in performance on the WAIS-R subtests could also provide a basis for further research to examine the extent to which each subtest measures formal reasoning or abstract thinking ability. This information could be useful in assessing the effectiveness of educational strategies developed to enhance formal reasoning ability in college students.

CHAPTER 2

Method

Subjects

The Arlin Test of Formal Reasoning (ATFR) and the Wechsler Adult Intelligence Test-Revised (WAIS-R) were administered to 20 subjects, 11 males and 9 females. The subjects ranged in age from 18 to 20 with a mean age of 19. The mean age for males in this sample was 18.8, and the mean age for females was 19.2. Subjects were college freshman or sophomore student volunteers who attended a midwestern university.

Prior to testing, each subject was required to read and sign a consent form which explained testing procedures, confidentiality issues, and the participants right to withdraw from the study at any time (see Appendix A). Subjects were asked to provide their age and gender on the test answer forms. A four digit number was assigned to each subject to eliminate the use of names or other identifying information of a personal nature in order to maintain confidentiality. Testing and data collection were initiated only after approval for this study was obtained from the Review Board of Human Subjects in accordance with university policies.

Instruments

The ATFR was the instrument chosen to measure formal reasoning ability. The ATFR uses a four response multiple-choice format and consists of 32 items. The reasoning problems are presented in printed form and are followed by appropriate response choices for each item. Items in the ATFR are also represented in graphic form by line drawings.

The 32 items have been organized into eight subtests which correspond to the eight formal concepts associated with formal reasoning ability. These concepts are: (a) multiplicative compensations, (b) correlations, (c) probability, (d) combinatorial reasoning, (e) proportional reasoning, (f) forms of conservation beyond direct verification, (g) mechanical equilibrium, and (h) the coordination of two or more systems of frames of reference.

The ATFR yields two sets of scores, a total score and eight subtest scores. The total raw score is used to assign cognitive level distinctions with five ranges, low concrete, high concrete, transitional, low format, and high formal. These levels are based on Inhelder and Piaget's (1958) description of performance (Arlin, 1984). The total score means for the ATFR range from 13.59 (<u>SD</u> = 4.31) for grades 8-6 to 18.33 (<u>SD</u> = 5.11) for grades 10-12.

A multitrait-multimethod validity study (Arlin, 1982) reported consistently high relationships in the comparison of correlation coefficients represented in the multitrait-multimethod matrix that were indicative of a high level of validity. The technique used to determine validity in this study is also consistent with recommendations based on a review by Nagy and Griffins (1982) citing criticisms of previous attempts to develop a valid paper and pencil group test for assessment of formal reasoning based on Piagetian tasks. Test-retest reliability ranged from .76 to .89 with the variation attributed to the use of different versions of the ATFR and the time period between test administrations. Therefore, information concerning the ATFR available in the multitrait-multimethod validity study (Arlin, 1982) and in the ATFR manual (Arlin, 1984) indicate the instrument is a valid method to assess formal operational ability.

The WAIS-R is the other instrument employed in this study. This instrument consists of 11 subtests, 6 verbal and 5 performance, which are designed to provide a forum for individuals to demonstrate a range of capabilities that can be evaluated in terms of scores obtained. These subtests are designed to measure different areas of mental abilities using a variety of methods.

The Verbal Scale of the WAIS-R is composed of the Information, Digit Span, Vocabulary, Arithmetic, Comprehension, and Similarities subtests. A brief summary of these scales follows. The <u>Information</u> subtest requires the subject to answer questions dealing with information accumulated in our society rather than information taught in a formal manner. <u>Digit Span</u> involves the subject's ability to repeat numbers in increasingly larger groupings both forward and backward; it measures memory, concentration and sequential processing abilities. Vocabulary requires the subject to define the words presented both visually and orally; it provides information concerning accumulated verbal ability, as well as provides insight into the thought processes. Arithmetic is an orally presented subtest that evaluates the ability to solve mentally mathematical problems involving numerical reasoning and the speed of numerical processing. Comprehension is a subtest that requires the subject to answer common sense questions measuring elements of practical judgement and self-direction. Similarities requires the subject to tell how two different things are alike; it is a measure of verbal abstract and conceptual thinking and is also used for evaluating the rigidity and flexibility of an individual's thinking.

The Performance Scale consists of the Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Digit Symbol subtests. Picture Completion requires the subject to tell what is missing from a succession of pictures; it measures general awareness of one's environment as well as the capacity for concentrated effort and visual conception. Picture Arrangement enjoins the subject to arrange a series of pictures in a sequence that tells a story that makes sense; it measures grasp of sequence, social planning, and comprehension of individual parts into a whole. Block Design involves colored blocks the subject manipulates to match printed designs; it measures visual motor coordination and perceptual organization. Object Assembly requires the correct arrangement of puzzle pieces; it provided measures of the ability to differentiate familiar configurations, to perceive relationships of unknown objects, and speed of manipulation. Digit Symbol involves copying marks from a code in the appropriate places; it is a measure of memory and retrieval of information, as well as attention span and distractibility.

Each subtest produces a raw score which is then converted to a scaled score according to a table of norms developed from normative samples. The sums of the scaled scores are converted to Performance and Verbal IQ scores. The Full Scale, Verbal and Performance Scales all have a mean of 100 (SD = 15), and the subtests each have a mean of 10 (SD = 3).

The WAIS-R is administered individually and takes approximately 60 to 90 minutes to complete. Reliability coefficients as reported in the WAIS-R manual (Wechsler, 1981) were computed for Full Scale IQ, with a range of .96 to .98, for Verbal IQ with a .95 to .97 range, and for the Performance IQ ranging from .88 to .94. Individual subtests coefficients ranged from .52 to .96. Variance in standard errors of measurement was also reported for each of the IQ subtest scores with error variance for Full Scale IQ below three, which indicated that an individual Full Scale score is accurate to a degree of plus of minus three points 68 times out of 100.

Although validity data in the WAIS-R manual is dependent on the research of a previous version of the Wechsler Adult Intelligence Scale (WAIS), summary information concerning the relevance of this data to the WAIS-R by Anastasi (1988) and Kaufman (1990) substantiate Wechsler's assumption that the WAIS-R is a valid measure of global intelligence. Consequently, the various components of research accessed by this author concerning the validity and reliability of the WAIS-R support the suitability of this instrument for use in this study.

Procedure

The Wechsler Adult Intelligence Scale-Revised and the Arlin Test of Formal Reasoning were administered to each subject. The WAIS-R was administered individually, and the ATFR was given in groups as well as individually depending on the scheduling needs and restrictions of each subject. Both tests were administered by the author according to the instructions provided in their respective manuals.

All tests completed by the subjects in this study were scored by the author. In recognition of the prevalence of examiner error in scoring of the WAIS-R by graduate students (Slate, Jones, Murray, & Coulter, 1993), the WAIS-R scores were reviewed by another individual whose training, course work, and experience with this instrument approximate that of the author.

Statistical Design

The statistical analyses of the data collected for this study consisted of comparing the ATFR scores with the Full Scale, Verbal, and Performance IQ scores of the WAIS-R as well as comparing the ATFR scores with each of the eleven subtest scores obtained from the WAIS-R. The Pearson product-moment correlation analyses, the most commonly used method of measuring strength of relationships between two variables (Shavelson, 1988), were used to determine the relationship between the ATFR scores with each of the above mentioned scores from the WAIS-R. A Fisher's table was used to establish significance.

In addition, the sample group was differentiated by gender. A series of *t*-tests were used to determine whether significant differences existed between scores on the tests with reference to gender differences. Means and standard deviations were also calculated to help describe the sample.

CHAPTER 3

Results

The three IQ scaled scores, Verbal, Performance, and Full Scale as well as the 11 subtests scaled scores of the Wechsler Adult Intelligence Scale-R (WAIS-R) were obtained from a sample group of college students ages 18 to 20. The 11 subtests of the WAIS-R include Information, Digit Span, Vocabulary, Arithmetic, Comprehension, Similarities, Picture Completion, Picture Arrangement, Block Design, Object Assembly, and Digit Symbol. The 14 scaled scores achieved by the sample group from the WAIS-R were each correlated with the total scores on the Arlin Test of Formal Reasoning obtained from the same sample group. The purpose of these comparisons is to examine the relationship between formal reasoning ability as measured by the ATFR to IQs, as measured by the WAIS-R.

The ATFR distinguishes scores in terms of five levels of cognitive reasoning, Concrete, High Concrete, Transitional, Low Formal, and High Formal. The number of students scoring at each of the five ATFR levels of reasoning was obtained and converted into percentages. These results are shown in Table 1.

Table 1

Percentage of Students Scoring at Each of the Five ATFR Levels of Reasoning

Levels of Reasoning	N	%
Concrete	2	10
High Concrete	1	5
Transitional	3	15
Low Formal	7	35
High Formal	7	35

N = 20.

The percentage of students in the sample group for this study scored in the formal reasoning range (Low Formal + High Formal) was 70.

Descriptive statistics were also run on the data collected for this study. The mean scores, standard deviations, and ranges obtained from the sample group of the total score of the ATFR and the three IQ scaled scores and 11 subtest scaled scores of the WAIS-R are shown in Table 2.

Table 2

Descriptive Statistics for ATFR and WAIS-R Scores

Test	*M*	*SD*	*R*
ATFR			
Total	21.50	6.74	6-31
WAIS-R IQ scales and subtests			
Full Scale IQ	114.25	14.08	96-142
Verbal IQ	109.45	12.61	90-134
Performance IQ	116.20	14.89	90-136
Information	10.15	2.16	6-14
Digit Span	9.10	2.68	6-16
Vocabulary	10.10	2.10	6-15
Arithmetic	10.50	1.86	7-13
Comprehension	11.30	2.74	6-14
Similarities	10.95	2.37	8-16
Picture Completion	10.45	2.01	7-14
Picture Arrangement	11.10	2.45	7-17
Block Design	12.25	3.29	6-15
Object Assembly	12.60	2.50	7-18
Digit Symbol	12.55	2.39	9-15

Note. ATFR = Arlin Test of Formal Reasoning, WAIS-R = Wechsler Adult Intelligence Scale-Revised.

The relationship between test scores was determined by using the Pearson product-moment correlation technique. These correlations are presented in Table 3.

Table 3

Correlations Between ATFR and the WAIS-R IQ Scales and Subtests

Tests	r
WAIS-R Full Scale IQ - ATFR	.58*
WAIS-R Verbal IQ - ATFR	.63*
WAIS-R Performance IQ - ATFR	.42
WAIS-R Information Subtest - ATFR	.75**
WAIS-R Digit Span Subtest - ATFR	.40
WAIS-R Vocabulary Subtest - ATFR	.67**
WAIS-R Arithmetic Subtest - ATFR	.27
WAIS-R Comprehension Subtest - ATFR	.30
WAIS-R Picture Completion - ATFR	.22
WAIS-R Similarities Subtest - ATFR	.41
WAIS-R Picture Arrangement Subtest - ATFR	.16
WAIS-R Block Design Subtest - ATFR	.52
WAIS-R Object Assembly Subtest - ATFR	.35
WAIS-R Digit Symbol Subtest - ATFR	.10

* $p > .01$

** $p > .001$

The correlational analyses of score data obtained from the Arlin Test of Formal Reasoning Total Score and the Wechsler Adult Intelligence Scale-Revised Full Scale IQ indicate a positive correlation exists between the two measures. The correlation coefficient between the these two measures was .58 ($p > .01$). The correlation between the WAIS-R Verbal IQ and the ATFR of .63 was also significant at the .01 level. Two of the WAIS-R Subtests, Information and Vocabulary, indicated a positive correlation with the ATFR with respective correlations of .75 and .65 ($p > .001$). The remaining WAIS-R IQ scale and subtest scores were not significantly correlated with the ATFR total scores.

In an effort to determine if any statistically significant differences in scores occurred between mean scores of male and female subjects on any of the test results used in this study, a series of *t*-tests was also calculated. Table 4 contains the results of mean difference tests for the subject pool when grouped on the basis of gender.

Table 4

Means, Standard Deviations, and t values of the ATFR and WAIS-R scales and subtests for Males and Females

	Males (N = 11)		Females (N = 9)		
Test	M	SD	M	SD	t
ATFR					
Total	12.18	8.15	21.11	4.99	.02
WAIS-R scales and subtests					
Full Scale IQ	117.18	14.46	110.66	13.54	1.04
Verbal IQ	111.27	13.10	107.22	12.37	.71
Performance IQ	120.18	14.42	111.33	14.77	1.35
Information	10.18	2.36	10.11	2.03	.07
Digit Span	11.00	2.32	8.77	2.28	2.16
Vocabulary	10.09	2.43	10.11	1.76	.02
Arithmetic	10.63	1.69	10.33	2.18	.34
Comprehension	9.45	3.01	11.44	3.24	1.41
Picture Completion	11.36	1.88	9.77	2.11	1.79
Similarity	11.36	2.73	10.44	1.88	.89
Picture Arrangement	12.27	2.19	9.66	2.00	2.78
Block Design	12.54	2.66	11.88	4.08	.42
Object Assembly	12.72	2.37	12.44	2.79	.24
Digit Symbol	12.72	2.83	12.33	1.87	.36

The obtained _t_ scores on all test scores from the ATFR and the WAIS-R used in this study were not greater than or equal to 2.861, the critical _t_ value at the .01 level of significance. Therefore, no significant differences between male and female group means were found on the ATFR or on any of the scores of the IQ scales and subtests of the WAIS-R.

CHAPTER 4

Discussion

According to Piaget's theory of cognitive development, the transition from concrete thinking to formal reasoning ability generally becomes apparent between the ages of 11 to 15 (Piaget & Inhelder, 1969). One aspect of this study was to determine the level of cognitive ability demonstrated by this sample group as measured by the Arlin Test of Formal Reasoning (ATFR).

The results of this study indicate 70% of the participants in this study demonstrated formal operational ability. Fifteen percent of the subjects in this sample were determined to be operating in a transitional status between concrete and formal reasoning ability with the remaining 15% shown to be operating at a concrete level of cognitive ability. This sample has shown a somewhat higher percentage of college students operating at the formal level of reasoning than other studies targeting 18 to 20 year old college students for similar purposes. Similar studies have shown as high as 52% of college students tested to be still operating in the concrete stage of cognitive development (Chiappetta, 1976; Primeau, 1989; & Logan, 1991). Concerns noted with respect to this research include the need to identify concrete thinkers at the college level in order to accommodate them through matching instructional methods and applications of formal reasoning strategies to facilitate academic achievement (Danner & Day, 1977; Bender & Milakofsky, 1982; Niaz, 1987).

Past research attempting to compare psychometric and cognitive developmental aspects of intelligence have indicated low to moderate relationships between the two (DeVries, 1974). The correlation of .58 for the comparison of the ATFR and the WAIS-R Full Scale IQ scores while statistically significant at the .01 level would suggest a moderate relationship does exist between these aspects of intelligence in this sample group. To more closely examine the elements of this relationship, the ATFR scores were also compared to the WAIS-R Verbal and Performance IQ scores of this sample group. The resulting correlations of .63 between the Verbal IQ and ATFR scores and .42 between the Performance IQ scores and the ATFR would indicate the components of the Verbal IQ test have a stronger relationship to formal reasoning ability as measured by the ATFR than do the components of the Performance IQ portion of the WAIS-R.

Further comparisons utilizing correlations of the total scores of the ATFR with each of the 11 subtests of the WAIS-R indicated only two subtests, Information and Vocabulary, as having statistically significant relationships to the ATFR. The Information and Vocabulary subtests are both component parts of the Verbal IQ scale of the WAIS-R. The correlations between these two subtests and the ATFR were .75 and .67 respectively, and both were statistically significant at the .001 level.

The relationships between ATFR and the Information and Vocabulary subtests in this sample group suggest the Information and Vocabulary subtests may have more potential to measure formal reasoning ability than do any other component parts of the WAIS-R. This would also suggest abstract or formal reasoning ability is a facet of intelligence that may be included at least in part in the psychometric process of measuring IQ.

In an effort to determine if differences in test scores obtained in this sample group could be related to differences in gender, the sample group was broken on a gender basis for additional analysis. A series of *t*-tests on mean differences for test scores obtained in this study showed no statistically significant differences. These results would tend to negate any relationships between differences in scoring to gender differences within the subject pool tested for this study.

It should be noted that the small sample size, limited by the accessibility to a larger sample population by time and financial restraints, should be taken into consideration when generalizing the results of this study. The timing of the data collection also may have some impact on the results as the tests were administered during academic summer sessions. Summer sessions may be a time when student populations are somewhat atypical of those student populations in fall and spring semesters, as enrollment is frequently much smaller and motivational factors for attendance may also be somewhat different. This factor is perhaps most significant as it pertains to the higher percentage of students operating at the formal reasoning level of cognitive development in this sample group than has been demonstrated in similar studies.

The statistically significant relationship between the Information and Vocabulary subtests of the WAIS-R and the ATFR in this sample group points to a need for further research to substantiate and further define such a relationship. Such research could provide information as to how to broaden the use of psychometric measures of IQ for use in conjunction with developmental aspects of intelligence thereby enhancing the capability to assess developmental levels of cognitive ability.

REFERENCES

Ahlawat, K. S., & Billeh, V. V. (1987). The comparative investigation of the psychometric properties of three tests of logical thinking. Journal of Research in Science Teaching, 24(2), 93-105.

Anastasi, A. (1988). Psychological testing (6th ed.) New York: Macmillan.

Arlin, P. K. (1975). Cognitive development in adulthood: A fifth stage? Developmental Psychology, 11, 602-606.

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. (1984). Arlin Test of Formal Reasoning: Test Manual. New York: Slosson Educational Publications.

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, 205-216.

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.

Brainerd, C. J. (1978). Piaget's theory of intelligence. Englewood Cliffs, NJ: Prentice Hall.

Chiappetta, E. L. (1976). A review of Piagetian studies relevant to science instruction at the secondary and college level. Science Education, 60, 253-261.

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

DeVries, R. (1974). Relationships among Piagetian IQ, and achievement assessments. Child Development, 45, 746-756.

Elkind, D. (1961). The development of quantitative thinking: A systematic replication of Piaget's studies. The Journal of Genetic Psychology, 98, 37-46.

Elkind, D. (1971). Two approaches to intelligence: Piagetian and psychometric. In D. R. Green, M. P. Ford, & G. B. Flamer (Eds.), Measurement and Piaget (pp.11-33). New York: McGraw Hill.

Evans, R. L. (1973). Jean Piaget the man and his ideas. New York: E. P. Dutton.

Evans, B., & Waites, B. (1981). IQ and mental testing: An unnatural science and its social history. Highlands, NJ: Humanities Press.

Fakouri, M. E. (1985). Arlin Test of Formal Reasoning. In D. J. Keyser, & R. C. Sweetland (Eds.), Test Critiques: Vol. 2. (pp.40-44). Kansas City, MO: Test Corporation of America.

Fancher, R. E. (1985). The intelligence men. New York: Norton.

Furst, E. J. (1950). Relationship between tests of intelligence and tests of critical thinking and knowledge. Journal of Educational Research, 43, 614-625.

Goddard, H. H. (1912). The Kallikak family: A study in the heredity of feeblemindedness. New York: MacMillan.

Guilford, J. P. (1936). Psychometric methods. New York: McGraw Hill.

Hacker, R. G. (1989). The construct validities of some group tests of intelligence and tests of critical thinking and knowledge. Journal of Educational Research, 43, 269-275.

Howe, M. J. A. (1989). Separate skills or general intelligence: The autonomy of human abilities. British Journal of Educational Psychology, 59, 531-360.

Hunt, T. (1936). Measurement in psychology. New York: Prentice-Hall.

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

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

Kaufman, A. S. (1990). Assessing adolescent and adult intelligence. Boston: Allyn and Bacon.

Kline, P. (1991). _Intelligence: The psychometric view_. New York: Rutledge.

Lawson, A. E. (1978). The development and validation of a classroom test of formal reasoning. _Journal of Research in Science Teaching_, _15_(1), 11-24

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.

Logan R. A. (1991). _Relationships between the Arlin Formal Reasoning and Classroom Multiple-Choice Tests_. Unpublished master's thesis. Emporia State University. Emporia, KS.

McKinnon J. W., & Renner, J. W. (1971). Are colleges concerned with intellectual development? _American Journal of Physics_, _39_, 1047-1052.

Nagy, P., & Griffiths, A. K. (1982). Limitations of recent research relating Piaget's theory to adolescent thought. _Review of Educational Research_, _52_, 513-556.

Niaz, M. (1987). The role of cognitive factors in the teaching of science. _Research in Science and Technological Education_, _5_, 6-16.

Piaget, J. (1972). Intellectual evolution from adolescents to adulthood. _Human Development_, _15_, 1-12.

Piaget, J., & Inhelder, B. (1969). _The psychology of the child_. New York: Basic Books.

Pratt, C., & Hacker, R. G. (1984). Is Lawson's Classroom Test of Formal Reasoning valid? _Educational and Psychological Measurement_, _44_(2), 441-448.

Primeau, C. K. (1989). _Relationship between scores on the Gorham Poverbs Test and the Arlin Test of Formal Reasoning_. Unpublished master's thesis. Emporia State University. Emporia, KS.

Raven, R. J. (1973). The development of a test of Piaget's logical operations. _Science Education_, _57_, 377-385.

Roberge, J. J., & Flexer, B. K. (1982). The Formal Operational Reasoning Test. _Journal of General Psychology_, _106_(1), 61-67.

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.

Santmire, T. E. (1985). Review of the Arlin Test of Formal Reasoning. In J. V. Mitchell, Jr. (Ed.) _The ninth mental measurement yearbook: vol 1_. (pp. 81-83). Lincoln, NE: University of Nebraska Press.

Shavelson, R. J. (1988). _Statistical reasoning for behavioral sciences_. Boston: Allyn and Bacon.

Shayer, M., Adey, P., & Wylam, H. (1981). Group tests of cognitive development: Ideals and realizations. _Journal of Research in Science Teaching_, _18_(2), 157-168.

Slate, J. R., Jones, C. H., Murray, R. A., & Coulter, C.

(1993). Evidence that practitioners err in administering and scoring the WAIS-R. Measurement and Evaluation in Counseling Development, 24,(4), 156-161.

Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. American Psychologist, 42, 137-144.

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

Stoddard, G. D. (1944). The meaning of intelligence. New York: Macmillan.

Strahan, D. B., & O'Sullivan R. G. (1988). Achievement test scores in middle grades: The influence of cognitive reasoning. Journal of Psychoeducational Assessment, 8, 53-61.

Terman, L. M. (1916). The measurement of intelligence. Boston: Riverside Press Cambridge.

Thorndike, R. M. (1990). Origins of intelligence and its measurement. Journal of Psychoeducational Assessment. 8, 223-230.

Tobin K. G., & Capie, W. (1981). The development and validation of a group test of logical thinking. Educational and Psychological Measurement, 41(2), 413-423.

Von Mayhauser, R. T. (1989). Making intelligence functional: Walter Dill Scott and applied psychological testing in

World War I. Journal of the History of Behavioral Sciences, 24, 60-72.

Walker, R. A., Hendrix, J. R., & Mertens, T. R. (1979). Written Piagetian task instrument: Its development and use. Science and Education, 63(2), 211-220.

Watkins, M. W. (1985). Review of WAIS-R a clinical report. In J. Mitchell (Ed.) The ninth mental measurement yearbook. Lincoln, NE: Buros Institute of Mental Measurements.

Wechsler, D. (1944). The measurement of adult intelligence. (3rd ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1958). The measurement and appraisal of adult intelligence. (4th ed.). Baltimore: Williams & Wilkins.

Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. American Psychologist, 30, 134-139.

Wechsler, D. (1981). Manual for the Wechsler Adult Intelligence Scale-Revised. New York: The Psychological Corporation.

Weinberg, R. A. (1989). Intelligence and IQ. American Psychologist, 44, 98-104.

Zachary, R. A. (1990). Wechsler intelligence scales: theoretical and practical considerations. Journal of Psychoeducational Assessment, 8, 276-289.

APPENDIX

INFORMED CONSENT DOCUMENT

The Department/Division of psychology supports the practice of protection for human subjects participating in research and related activities. The following information is provided so that you can decide whether you wish to participate in the present study. You should be aware that even if you agree to participate you are free to withdraw at any time, and that if you do withdraw from the study, you will not be subjected to reprimand or other form of reproach.

Procedures to be followed in the study, as well as identification of any procedures which are experimental.

Participants in this study will take two tests, the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and the Arlin Test of Formal Reasoning (ATFR). The WAIS-R will be administered to each subject individually, and the ATFR will be given in groups although individual administration will be considered if scheduling conflicts dictate the need to do so. Both tests will be administered by Glenda Young according to instructions provided in the respective manuals for each instrument. A numbering system will be used to identify students to insure confidentiality. Date of birth and gender will be the only information recorded for each participant.

Description of any attendant discomforts or other forms of risk involved for subjects taking part in the study.

Any information obtained through the administration of either test will be kept confidential.

Descriptions of benefits to be expected from this study or research.

The information obtained from this study will be helpful in exploring the relationships between the level of cognitive development achieved and IQ. The scope and nature of the relationship(s) identified in this study are intended to provide a basis for further research which could be useful in assessing the effectiveness of educational strategies developed to enhance formal reasoning ability in college students.

I have read the above statement and have been fully advised of the procedures to be used in this project. I have been given sufficient opportunity to ask questions I had concerning the procedures and possible risks involved. I understand the potential risks involved and assume them voluntarily. I likewise understand that I can withdraw from this study at any time without being subjected to reproach.

______________________________________ _____________

subject and/or authorized representative date

TO: All Graduate Students Who Submit a Thesis or Research/Project as Partial Fulfillment of The Requirements for an Advanced Degree.

FROM: Emporia State University Graduate School

I, Glenda Young, hereby submit this thesis/report to Emporia State University as partial fulfillment of the requirements for an advanced degree that the Library of the University may make it available for use in accordance with its regulations governing materials of this type. I further agree that quoting, photocopying, or other reproduction of this document is allowed for private study, scholarship (including teaching) and research purposes of a nonprofit nature. No copying which involves potential financial gain will be allowed without written permission of the author.

Glenda Young
Signature of Author

11-29-93
Date

Relationships Between the Online Test of Formal Reasoning and the Wechsler Adult Intelligence Scale - Revised
Title of Thesis/Research Project

Signature of Graduate Office Staff Member

11-29-93
Date Received

Distribution: Director, William Allen White Library
Graduate School Office
Author