AN ABSTRACT OF THE THESIS OF

Matthew S. Schrader for the Master of Science

in Psychology presented on December 13, 1994

Title: The Test-Retest Reliability of the WAIS-R For the 18 and 19 Year Age Group

Abstract Approved: Cooper B. Holmes

In a national survey, Piotrowski and Keller (1989) found the Wechsler Adult Intelligence Scale - Revised (WAIS-R) was the second most popular psychological and/or educational test administered to adults. Information derived from the scales has many uses, including placement in schools, clinical diagnosis, and assessing neuropsychological damage. Therefore, the WAIS-R's reliability in all age groups is crucial.

One method of assessing a test's stability is administering the same test twice with a specified time interval and then correlating the scores from the two testings. Wechsler (1981) published normative data in the WAIS-R manual for the retest reliability of only the 25 to 34 age group and the 45 to 54 age group, leaving seven age groups unexamined, including the 18 and 19 age group. Wechsler suggested further research be done on groups of differing ability levels. Previous research has suggested the WAIS-R's stability is suspect in the under 20 age groups. The present study was designed to investigate the retest stability and the magnitude of practice effects evidenced when the WAIS-R was utilized in the 18 and 19 age group. Wechsler (1981) exerted considerable effort to insure no gender bias existed in the WAIS-R; however, gender differences have been noted in several studies, therefore, the effect of gender on gains in IQ scores was examined in this study.

In the current study, 44 18 and 19 year-old college students (15 males, 29 females) who were enrolled in undergraduate classes, were tested twice within a two-month period, similar to the process employed by Wechsler. Retest correlations ranged from .62 to .89 for the combined sample on the eleven subtests and were .90, .88, and .94 for the Verbal, Performance, and Full Scale IQs, respectively. Furthermore, the men gained significantly more on Verbal and Full Scale IQs from test to retest.

THE TEST-RETEST RELIABILITY OF THE WAIS-R FOR THE 18 AND 19 YEAR AGE GROUP

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

Matthew S. Schrader

December 13, 1994

Kenneth A. Weaver

Approved for the Division of Psychology
and Special Education

John E. Schuerman

Approved for the Graduate Council

ACKNOWLEDGMENTS

I would like to convey my sincere gratitude to my thesis committee chairman and faculty advisor, Dr. Cooper B. Holmes. His guidance and friendship throughout my graduate career have been greatly appreciated. I would also like to thank Dr. Phillip Wurtz for his assistance as a committee member for this thesis. I would like to take this opportunity to extend my heartfelt thanks to Mr. Howard Carvajal. His attention to detail, dedication, and knowledge of the Wechsler scales is truly astounding and without his advice, this thesis would not have been possible. Thanks are also in order for Heather Kirchhefer for her assistance in collecting data. Finally, I would like to thank all the college freshmen and sophomores who took the time out of their busy schedules to enable the completion of this thesis.

A special thanks goes to my parents for making my academic endeavors possible. Their understanding and support throughout the years has been invaluable. Finally, I express my deepest gratitude to Ami, my fiancée, for her continual encouragement, understanding, and love.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

Introduction

The Wechsler intelligence scales are currently the most widely used measures of intelligence in a variety of professional settings (Lubin, Larsen, & Matarazzo, 1984). In fact, in a national survey, Piotrowski and Keller (1989) discovered the WAIS-R was the second most popular psychological and/or educational test administered to adults. The Minnesota Multiphasic Personality Inventory was the most commonly utilized test. Information derived from the scales has many uses, including placement in schools, clinical diagnosis, and assessment of neuropsychological damage (Kaufman, 1990). Therefore, the WAIS-R's reliability across all age groups is crucial.

For the purpose of this thesis and because there is no universally accepted definition of intelligence, Wechsler's (1958) definition will be used: "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" (p. 7). Wechsler does not define intelligence as a single entity but an overall ability to comprehend effectively and deal with one's environment. Therefore, intelligence is a function of the personality as a whole (Wechsler, 1981).

Equally as important as establishing an appropriate definition of intelligence is discussing the general development of adult intellectual functioning. The growth of intellectual functioning or mental age is generally depicted as a curve that gradually increases from birth to late adolescence and early adulthood (Bayley, 1955; Wechsler, 1950). An individual's IQ can be compared with the mean IQ of individuals of the same

age since one's level of intelligence remains relatively constant across one's lifespan (Klonoff, 1972).

Although the WAIS-R has been extensively studied, the retest reliability of the WAIS-R for certain age groups still is not known. Retest reliability refers to a test's yielding the same results after repeated administrations to the same participant (Bootzin, Acocella & Alloy, 1993). Naglieri and Pfeiffer (as cited in Watkins & Campbell, 1992) stated that reliability and stability are important because they indicate the likelihood and degree to which an individual's scores vary over time. The retest reliability for the 18 and 19 age group is not presented in the WAIS-R manual although data for the 25 to 34 and 45 to 54 year old age groups show high coefficients of stability (Wechsler, 1981). Indeed, the stability of IQ scores in the 18 and 19 age group is not mentioned in the current literature.

Thompson and Molly (1993) asserted that "test-retest base rate data need to be developed for different samples and . . . routinely considered by WAIS-R practitioners" (p. 891). Matarazzo and Herman (1984) have remarked:

> Research showing the stability of the Wechsler (or other neuropsychological test) scores obtained from test to retest for a single individual has been relatively scarce. Given the costs, both human and financial, involved for the examiner and examinee, the dearth of such retest studies is not surprising. Nevertheless, the accelerating increase in the frequency of retests of the same individual using the Wechsler scales by clinical neuropsychologists requires that certain base-rate information relative to

> the Wechsler scales be available to these practitioners. The need for such base rates, hopefully derived over time from many clinical as well as normal samples of individuals, becomes clear to a reader from the differences in perceptions one will discern in the differing interpretations offered by investigators who have addressed the meanings of the same test-retest changes in WAIS scores (p. 351).

Matarazzo and Herman (1984) stated further:

> if the practice of psychology is to be built upon increasingly firm scaffolding, actuarial base rates should be developed from different samples, of sufficient size, which differ on such potentially important variables such as age and the other variables [e.g., race, level of intellectual functioning, socioeconomic status] identified by Shatz and other writers (p. 359).

As cited in Slate, Frost, & Cross (1991), Gregory stated "the WAIS-R appears to have 'peculiar norms' for persons below 20 years of age; [sic] norms which may not be representative of the general population" (p. 5). Slate et al.'s (1991) findings corroborate the above observation and support Kaufman's assertion (cited in Slate et al., 1991) that the validity of WAIS-R test data for persons under the age of 20 is dubious at best.

Wechsler himself stated "further studies are needed to establish the magnitude of practice effects at various levels of ability" (Wechsler, 1981, p. 31). Even though Wechsler does not specifically name the 18 and 19 age group as needing investigation of practice effects, this age group is presumably at a unique level of ability as deduced from their meriting their own IQ tables. Such endeavor would serve would also give clinicians

who employ the WAIS-R base rate data that could serve as a "yardstick" against which to measure their own patients' retest scores. Establishing the reliability of this widely used psychometric tool is an initial and essential step in the validation process. Proven reliability is a crucial characteristic for the clinical application of any psychometric tool (Kowalski & Rossini, 1990). Given this statement and the aforementioned dearth of literature on the subject, Wechsler's recommendation to investigate the retest reliability in this specific age group seems reasonable.

Stability of Intelligence

As previously noted, Wechsler studied the effect the retest procedure had on the Verbal, Performance, and Full Scale IQ scores for the 11 subtests and 3 IQs for the 25 to 34 and the 45 to 54 age groups. These coefficients were determined by administering the WAIS-R twice (with two to seven week intervals). The stability coefficients for the 25 to 34 age group were .94 for Verbal IQ, .89 for Performance IQ, and .95 for Full Scale IQ. For the 45 to 54 age group, they were .97 for Verbal IQ, .90 for Performance IQ, and .96 for Full Scale IQ. The coefficients were determined by retesting a sample of 71 individuals 25 to 34 year olds and a sample of 48 45 to 54 year olds (Wechsler, 1981).

Stability coefficients obtained by Wechsler are an example of psychometric reliability, that is, scores characterized by high retest coefficients. Ryan, Georgemiller, Geisser, and Randall (1985) have noted that high clinical reliability is demonstrated by the absence of a meaningful change in score from initial test to retest. Ryan et al. (1985) retested a sample of 21 psychiatric and neurological patients after 2 to 144 weeks and

reported satisfactory psychometric reliability (e.g., $\underline{r} = .84$), but clinical reliability was unimpressive. For example, 86% of the patients' Full Scale IQ changes exceeded the standard error of measurement reported by Wechsler (1981) and 38% actually changed (26% up; 12% down) IQ classifications . These results insinuate the WAIS-R is less stable than one might deduce from the retest coefficient alone and large changes in IQ on retest must be interpreted in conjunction with information from medical, social, and educational sources (Ryan et al., 1985).

Atkinson et al. (1990) designed a study to assess the stability of WAIS-R factor scores over time in a sample of 39 mentally retarded (IQs < 80) individuals and 17 participants who evidenced heterogeneous IQ scores. They found impressive retest stability for both samples. The results revealed, in contrast to Ryan et al.'s 1985 study, 38 (97%) of the mentally retarded clients and 14 (82%) of the heterogeneous IQ participants did not show changes beyond what would be attributable to measurement error. Atkinson et al. (1990) states these results lend further credence to WAIS-R retest reliability.

Watkins and Campbell (1992) administered the WAIS-R to 50 mentally retarded adults on two separate occasions with a mean retest interval of two years, eight months. All participants had obtained WAIS-R Full Scale IQs of 69 or less. The mean ages for the two testings were 30.72 years and 33.52 years, respectively. Watkins and Campbell's results indicated that the WAIS-R possessed generally good stability overall although the retest reliability coefficients were quite variable. Some subtests such as Block Design showed good reliability over time ($\underline{r} = .82$); however, others such as Comprehension did

not (r = .17). A possible explanation offered for these results is that all subtests are clinically reliable (i.e., the absence of a meaningful change in score from test to retest) but in some instances are psychometrically unstable (i.e., scores characterized by low retest coefficients). The summary IQ data proved to have better reliability with retest correlations all in the .70s or .80s; thus, the authors concluded the WAIS-R was reliable for mentally retarded adults.

Matarazzo and Herman (1984), in investigating the Verbal IQ - Performance IQ differences in the WAIS-R standardization data for the 119 individuals who were retested, found that 24 (20.2% of the sample) participants decreased in Verbal IQ, 10 (8.4%) in Performance IQ, and 8 (6.7%) in Full Scale IQ. They pointed out that even though very few participants showed a decrease in IQ scores, any loss is significant for the practitioner given the "practice effects" shown by the group as a whole that produced higher rather than lower retest scores.

Moore et al. (1990) conducted a study utilizing 48 male and 12 female head-injured patients with a mean test-retest interval of 8.48 months. Their results were consistent with previous research in finding highly significant, positive retest stability coefficients that mirror those found in Wechsler's normative sample. Moore et al. (1990) noted that head-injured patients appear to produce larger retest changes in both WAIS-R IQs and subscale scores when evaluating their results in comparison to more general normative samples.

Paolo and Ryan (1993) conducted a study evaluating the retest reliability of the Satz-Mogel WAIS-R Short Form in a sample of 61 people 75 years and older. They

found that retest correlations for the Verbal ($\underline{r}$ = .80 to .82) and Full Scale ($\underline{r}$ = .79 to .83) IQs were for the most part acceptable but that the stability coefficients for the Performance IQ were unacceptable ($\underline{r}$ = .58 to .70); however, the validity and reliability of selected item short forms have traditionally been questioned (Kaufman, 1990). In contrast to Wechsler's standardization data (e.g., Object Assembly $\underline{r}$ = .68), Boone's (1992) study involving 100 psychiatric inpatients found high reliabilities for composite scales and subtests except for the Object Assembly subtest ($\underline{r}$ = .38). Boone (1992) suggests that Object Assembly's poor reliability and large standard error of estimate (2.22) could have been due to chance and cautions interpreting retest gains on Object Assembly with psychiatric inpatients. Concurrent with Matarazzo and Herman's (1984) and Kaufman's (1990) findings, the subtests with the lowest reliabilities evidenced the greatest improvement when the WAIS-R was readministered.

Schuerger and Witt (1989) analyzed retest reliability data from 34 separate studies to evaluate the impact of the partipant's age and status, the interval between test and retest, and gender of the participants. The data were obtained from the Stanford-Binet (third edition), the Wechsler Intelligence Scale for Children (WISC), the Wechsler Intelligence Scale for Children - Revised (WISC-R), the Wechsler Adult Intelligence Scale (WAIS), and the WAIS-R. The results indicated that the retest reliability was related strongly to the length of the interval and age but was not influenced by patient or normal status. In addition, once the effects of age and interval were accounted for, the instruments did not differ in retest reliability. Kowalski and Rossini (1990) computed the factor reliability

coefficients at each age level of the standardization sample. Their analysis revealed that the two primary factors Verbal Comprehension and Perceptual Organization had high reliability at all age levels (mean $\underline{r}$ = .96 and .90, respectively). The third factor Freedom from Distractibility also exhibited good reliability (mean $\underline{r}$ = .90) at all age levels.

Quereshi and Ostrowski (1985) conducted a study in which three Wechsler adult intelligence scales (the WAIS-R, WAIS, and Wechsler Bellevue II) were administered to 72 randomly selected undergraduates between the ages of 18 and 23. Results suggested the three intelligence scales were not parallel; however, the significance of Quereshi and Ostrowski's data to the present study is that the practice effect was not significant for any of the six verbal subtests, but it was significant for all of the performance subtests. Among the summary IQs, Verbal was not influenced by practice effects, Performance was highly affected by practice effects, and Full Scale was moderately affected by practice effects. Since the majority of the aforementioned studies demonstrate the relatively high stability of the WAIS-R in a variety of samples, it was hypothesized the stability for the 18 and 19 age group would be commensurate with those found for the other populations studied.

Snow, Tierney, Zorzitto, Fisher, and Reid (1989) examined the retest stability of the WAIS-R Verbal, Performance, and Full Scale IQs in a sample of 101 older normal individuals (mean age = 67.1). Retest reliability (Pearson $\underline{r}$) over this period was .86 for Verbal IQ, .85 for Performance IQ, and .90 for Full Scale IQ. The authors concluded for normal, healthy, elderly individuals, the reliability of the three summary IQ scores was high over a one year period. Snow et al. (1989) also found the majority of the subtest scaled

scores were stable over the one year period, although these correlations were lower than those for the summary IQ scores.

Effect of Gender on the Stability of Intelligence

Wechsler exerted considerable effort to insure no gender bias existed in the WAIS-R. However, sex differences on the Wechsler tests have been reported in the literature. The discussion of sex differences on the Wechsler tests typically focuses on performance on global measures of intelligence, specific subtests, or individual items and/or items whose content contains more references to one sex than the other (Ilai & Willerman, 1989).

Zoref and Williams (as cited in Ilai & Willerman, 1989) analyzed the content prejudice of numerous IQ scales, including the WAIS. They concluded there was "an overwhelming sexual and racial imbalance in the item content of IQ tests" (p. 226). The study revealed that of all WAIS items that made any reference to sex, 87% referred to males, while 13% referred to females. Ilai and Willerman (1989) conducted a study consisting of 206 Caucasian adults (110 males, 96 females). The men averaged 19.6 (SD = 3.3, range = 16 to 32) years of age and 12.9 years of school, while the women averaged 19.8 (SD = 3.2, range = 16 to 31) years of age and 12.8 years of school. The average Full Scale IQ for males was 110.9 (SD = 13.5), and for females 107.2 (SD = 12.6). These significantly ($p < .05$) different means were equivalent to sex differences noted by Matarazzo (1972, p. 353). The mean Verbal IQs for males of 109.2 (SD = 13.2) and for

females of 103.6 (SD = 12.7) were significantly different. The mean gender difference for Performance IQ was not significant.

Other sources (Kaufman, McLean, & Reynolds, 1988; Reynolds, Chastain, Kaufman, & McLean, 1987) corroborate that gender has an effect on performance on specific subtests. Men were superior to women on Information, Arithmetic, Picture Completion, Picture Arrangement, and Block Design, whereas women were superior to men on Digit Symbol. Overall, males earn higher mean scores than females by approximately 2 points on Verbal IQ, 1.5 points on Performance IQ, and 2 points on Full Scale IQ (Reynolds et al., 1987).

Psychologists in the People's Republic of China revised and standardized the WAIS recently for clinical, educational, and research interests in their country. The Chinese modified seven subtests (Dai, Gong, & Zhong, 1990, as cited in Dai, Ryan, Paolo, & Harrington, 1991) into the Wechsler Adult Intelligence Scale-Revised for China (WAIS-RC). Its overall factor structure and variance distribution are very similar to those of the WAIS and the WAIS-R (Dai et al., 1991). Dai et al. (1991) were the first to conduct a study on gender differences in China using the WAIS-RC. Overall results were comparable to those for American participants in the WAIS-R standardization sample (Reynolds et al., 1987). As in the American standardization sample, the Chinese men significantly outscored the Chinese women on the majority of the subtests and the three summary IQs, while Chinese women performed better than Chinese men on the Digit Symbol subtest (Dai et al., 1991).

Further evidence supports the contention that the WAIS-R has differing degrees of reliability in various age groups. Kaufman (1990), based on his research, has concluded that the WAIS-R norms for 16- to 19-year olds are suspiciously low, possibly due to a sampling bias, and are likely to produce falsely inflated IQs by three to five points. Thompson and Molly (1993) assessed the stability of the Wechsler Adult Intelligence Scale-Revised with a sample of 52 students who were 16 years of age. Half of the participants were retested after a 3-month interval and half after an 18-month interval. Males and females in the 18- relative to the 3- month retest interval group significantly increased Verbal IQ. Males in the 18- relative to the 3- month retest interval group increased more on Performance IQ. Females evidenced no short (3-month) versus long interval (18-month) retest discrepancy on Performance IQ. Males in the 18-month interval retest category increased significantly more than males in the 3-month interval retest group on Full Scale IQ; however, the short interval (3-month) versus long interval (18-month) retest increases for females were not significantly different.

In search of an additional indicator of intelligence, biological differences between men's and women's brains have been investigated by various researchers (Ankney, 1992; Kaufman, McLean, & Reynolds, 1991; Willerman, Schultz, Rutledge, and Bigler, 1992). Women's brains are smaller than men's, even after accounting for body size (Ankney, 1992). However, Jerison (as cited in Willerman et al., 1992) has reported more convoluted brains in women (e.g., more cortical folding) and thus more cortical surface than men. Haug (as cited in Willerman et al., 1992) suggested that women have greater

cortical neuron density than men. Taken together, these findings may mean that cortical surface area rather than brain size might be a better predictor of intelligence in women and men (Willerman et al., 1992). The Willerman et al. (1992) study was composed of 39 healthy college students whose MRI-derived hemispheric size differences were correlated with prorated WAIS-R Verbal IQ minus Performance IQ and Vocabulary minus Block Design scores within participants. The results of this study indicated:

> The pattern of women's cognitive strengths and weaknesses is dependent on relative hemisphere size as it is in men, but the pattern is somewhat different. The results for the men are consistent with the idea that left hemisphere size is more predictive of verbal function whereas right hemisphere size is more predictive of nonverbal function. The reversed-hemisphere size-ability relations in the women suggest that neural structures underlying their nonverbal problem solving are distributed to both hemispheres (Willerman et al., 1992, p. 324).

Purpose of Study

The purpose of this study was to assess the retest reliability of the Wechsler Adult Intelligence Scale - Revised (WAIS-R) with the 18 and 19 year-old age group. The rationale of this study was to assess the magnitude of practice effects on the reliability of the WAIS-R and establish the stability of one of the normed age levels. It was hypothesized in the present study that the 18 and 19 year old participants will show retest gains on the WAIS-R that are proportionate to the gains established in the 25 to 34 (e.g., 3, 9, and 7 points for Verbal, Performance, and Full Scale IQs, respectively) and 45 to 54

year old age groups (e.g., 3, 8, and 6 points for Verbal, Performance, and Full Scale IQs, respectively). Furthermore, it was postulated gender effects on the stability of intelligence will be congruent with the gender effects that have been shown in previous studies, that is, men will do better on Full Scale and Verbal IQs and no difference will emerge on Performance IQ.

CHAPTER 2

Method

Participants

The sample for this study consisted of 15 male and 29 female 18 and 19 year-old undergraduate volunteers. The participants were obtained from classes at a midwestern university. Additional credit toward the class grade was requested on behalf of the participants in return for their participation.

Testing Instrument

The Wechsler Adult Intelligence Scale - Revised (WAIS-R), which was published in 1981, was used to obtain the scores of each participant. The WAIS-R's development extended directly from its predecessors, the Wechsler Adult Intelligence Scale (WAIS), published in 1955, and the Wechsler-Bellevue Intelligence Scale, first published in 1939. The revision of the WAIS was undertaken to ensure its continued effectiveness as a basic test of adult intelligence and as a valid diagnostic tool and research instrument. The primary objective of the Wechsler Adult Intelligence Scale - Revised was to update the WAIS content and to provide new norms based on the responses and scores obtained from contemporaneous samples of the population. The WAIS-R retained much of the content and many of the features of the WAIS. The main changes involved revising or dropping some items that appeared dated and adding new ones. Altogether, about 80% of the WAIS-R items were retained from the WAIS, either unaltered or with only slight modifications. Other selected items were modified to reflect changes in item difficulty and

scoring to keep pace with advances in data analyses (Wechsler, 1981). Wechsler used as a norming sample groups that were considered representative of the adult population of the United States. The participant's ages ranged from 16 years, 0 months to 74 years, 11 months. A stratified sampling plan was utilized that was designed to include representative proportions of the United States general population. The 1970 United States Census and succeeding population reports furnished the basis for stratification upon the following variables: age, sex, race (white-nonwhite), geographic region, occupation, education, and urban-rural residence.

The WAIS-R consists of 11 subtests that are divided into Verbal and Performance categories. Wechsler intended that each subtest should correlate with Full Scale intelligence, that the subtests should test a variety of functions although they should not penalize persons with special abilities or disabilities, and that the responses to the items and subtests should have some diagnostic implications. Verbal and Performance subtests are administered in alternating order. The Verbal subtests include Information, Digit Span, Vocabulary, Arithmetic, Comprehension, and Similarities. A brief summary of these scales follows. The Information subtest measures accumulated information which has not specifically been taught and alertness to the everyday world. Combined with Vocabulary, one has an excellent test of general information, which correlates highly with Full Scale IQ. The Comprehension subtest measures common sense, the ability to evaluate past experience, and judgment in practical (social) situations. The Arithmetic subtest measures the ability to learn and retain mathematical material which was taught and to utilize that

which was learned. In addition, it measures the capacity for concentration and freedom from anxiety and distractibility. The Similarities subtest measures generalizing, categorical thinking, and verbal abstract conceptual thinking. The Digit Span subtest measures attention, concentration, freedom from distractibility, and/or immediate auditory recall (not delayed memory). The Vocabulary subtest quantifies accumulated verbal learning and the quality and character of thought processes. In addition, this subtest measures long-term cumulative learning and the ability to learn and retain concepts, associate them with concrete or abstract ideas, and recall them when necessary.

The Performance section is composed of Picture Completion, Picture Arrangement, Block Design, Object Assembly and Digit Symbol. The Picture Completion subtest measures the individual's ability to observe one's surroundings or environment with sufficient awareness to identify any details that might be missing. High scores on this subtest may indicate good perception, concentration, and interest in the environment. The Picture Arrangement subtest measures grasp of sequence, social planning and judgment, the ability to anticipate, and the ability to comprehend a total situation. The Block Design technique measures nonverbal reasoning, manipulative and perceptual speed, visual motor coordination, perceptual organization, and capacity for sustained effort. Speed contributes to a high score because of bonus points for quick arrangements. Object Assembly measures the ability to differentiate configurations, to perceive relationships of unknown objects, manipulative and perceptual speed, and visual-motor coordination.

Bonus points are available for fast, accurate assemblies. The Digit Symbol subtest measures visual rote learning and attention to detail.

The deviation IQ scores obtained by an individual are compared with the mean scores of the reference group that was included in the standardization of the WAIS-R. The mean score for Verbal, Performance and Full Scale IQs is 100 with a standard deviation of 15. The subtest standard scores are presented in a similar fashion. The mean score of any given subtest is 10 with a standard deviation of three. Verbal IQ is derived by summing the Verbal subtest scale scores and then comparing this sum to a scale that gives one a corresponding IQ score for the Verbal subtests. Performance IQ is found by the same procedure as Verbal IQ except the sum of the Performance subtest scale scores are used in obtaining an IQ score. Full Scale IQ is calculated by summing the totals of the Verbal and Performance IQ subtest scores and then using a table that converts the scale score points into IQ points. Full Scale IQ is the most commonly utilized score when discussing level of intelligence.

Procedure

Preceding the testing, the participants were presented with and asked to sign consent forms informing the participants of the testing procedure, confidentiality issues, and their right to withdraw from the proposed study at any time. Gender and age information were included; however, no names or other personal data were used or recorded in this study. Participants were assigned a number that was henceforth utilized when discussing or reporting scores in order to guarantee confidentiality.

Permission to conduct the study was obtained from thesis committee members and from the Institutional Review Board for Treatment of Human Subjects. The participants were tested by the author and fellow graduate students who had satisfactorily completed PY841 Clinical Mental Tests II (the Wechsler scales). The WAIS-R was administered twice to each volunteer. More specifically, this study consisted of volunteers who were healthy, 18 and 19 year-old college students. WAIS-R test and retest scores from this group were obtained and compared on the basis of Verbal, Performance, and Full Scale IQ scores, as well as on the scores of each of the 11 subtests that comprise the WAIS-R. The retest was completed within two to eight weeks, similar to the procedure used by Wechsler with the 25-34 and 45-54 age groups as noted in the manual (Wechsler, 1981). The tests were given during the fall semester of 1994 in rooms designed specifically for testing. The tests were scored by persons who have successfully completed the testing course mentioned above. The scoring was reviewed by a fellow graduate student who also had successfully completed PY841 Clinical Mental Tests II (the Wechsler scales) and who was uninvolved in the present study.

CHAPTER 3

Results

The present study was designed to assess the retest stability of the Wechsler Adult Intelligence Scale - Revised (WAIS-R) on 18 and 19 year-old college students. The mean time interval between test and retest was 23.9 days with a range of 14 to 63 days. The Pearson correlation was used to establish the stability of all 14 WAIS-R scores, IQs and subtest scores. The means and standard deviations of all scores for both test and retest were calculated and compared statistically. For stability coefficients for all 14 WAIS-R scale and IQ scores, see Table 1. The Digit Span's correlation coefficient for the combined sample as obtained in this study was .79. The correlation coefficient obtained by Wechsler for Digit Span was .89 for the 25 to 34 year old age group and .82 for the 45 to 54 year old age group. Comprehension's correlation coefficient in the present study was .68, the correlation obtained by Wechsler was .79 for the 25 to 34 year old age group and .82 for the 45 to 54 year old age group. Similarities evidenced the lowest stability of any subtest in the current study at $r = .62$. Wechsler's normative data show Similarities as having a correlation coefficient of .82 in the 25 to 34 year old age group and .86 in the 45 to 54 year old age group.

The Picture Completion subtest had a Pearson r of .76 in the present study while Wechsler's published norms show a Pearson r of .86 for the 25-34 year old age group and .89 for the 45-54 year old age group. In addition the Verbal IQ correlation coefficient

Table 1

Stability Coefficients of WAIS-R Subtests and IQs

Test	Men (N=15)	Women (N=29)	Combined Sample (N=44)
	r	r	r
Information	.96	.81	.89
Digit Span	.85	.76	.79
Vocabulary	.95	.83	.87
Arithmetic	.87	.75	.80
Comprehension	.82	.61	.68
Similarities	.75	.56	.62
Picture Completion	.86	.73	.76
Picture Arrangement	.85	.66	.72
Block Design	.81	.85	.82
Object Assembly	.87	.67	.74
Digit Symbol	.71	.89	.82
Verbal IQ	.97	.85	.90
Performance IQ	.91	.86	.88
Full Scale IQ	.97	.93	.94

Note. All correlations greater than .393 are significant at $p < .01$

obtained was .90 for the combined sample. Wechsler found a Verbal IQ correlation of .94 for the 25 to 34 year old age group and .97 for the 45 to 54 year old age group.

Men and women obtained Verbal IQ correlations of .97 and .85 respectively, from test to retest. Men achieved a Performance IQ correlation of .91 versus the women's .86. The men's correlation for Full Scale IQ was .97 while women evidenced a Pearson r of .93 from test to retest. For means and standard deviations by gender and combined sample, see Tables 2 and 3.

Three separate analyses of covariance were performed on the Full Scale, Verbal, and Performance IQs. Gender was the independent variable and the first testing score was the covariate. The statistical analysis of the Full Scale IQs yielded significance, $F(1, 41) = 4.05$, $p < .06$. The men's relative to women's Full Scale IQ was higher after retesting. The ANCOVA also showed significance for the Verbal IQs, $F(1, 41) = 9.30$, $p < .01$. The men's relative to women's Verbal IQ was higher after retesting. Performance IQs did not significantly differ, $F(1, 41) = .01$, $p < .05$ (see Tables 4, 5, and 6, respectively).

Table 2

Means and Standard Deviations of WAIS-R Subtests and IQs (By Gender)

	Males (N=15)				Females (N=29)			
	First Testing		Second Testing		First Testing		Second Testing	
Test	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Information	9.3	2.6	9.9	2.6	9.0	1.6	9.2	1.2
Digit Span	9.9	2.1	10.7	2.1	9.9	1.9	10.3	2.1
Vocabulary	9.9	2.5	10.2	2.5	10.0	1.7	9.5	1.3
Arithmetic	9.9	2.9	10.7	2.6	9.5	2.1	10.3	2.4
Comprehension	10.4	2.5	10.8	2.7	10.7	2.4	10.3	2.0
Similarities	10.7	2.5	12.3	1.9	10.4	2.1	11.1	1.8
Picture Completion	9.7	1.8	10.7	1.8	9.3	2.3	10.5	2.2
Picture Arrangement	10.0	3.1	10.7	2.7	10.1	2.8	11.2	3.0
Block Design	11.1	2.3	12.9	2.1	11.1	2.6	11.7	2.9
Object Assembly	11.1	3.7	13.7	3.1	11.1	3.2	13.5	2.8
Digit Symbol	12.3	2.0	12.7	1.8	12.9	2.6	14.3	2.1
Verbal IQ	107.4	13.7	112.6	11.9	106.3	7.1	108.1	7.3
Performance IQ	109.0	13.7	120.3	12.5	109.6	17.0	120.5	14.8
Full Scale IQ	109.5	13.8	117.7	11.7	109.0	11.4	114.9	10.9

Table 3

Means and Standard Deviations of WAIS-R Subtests and IQs (Combined Sample)

Combined Sample

(N=44)

	First Testing		Second Testing	
Test	Mean	SD	Mean	SD
Information	9.1	2.0	9.5	1.8
Digit Span	9.9	2.0	10.5	2.1
Vocabulary	10.0	2.0	9.7	1.8
Arithmetic	9.7	2.4	10.4	2.5
Comprehension	10.6	2.4	10.5	2.3
Similarities	10.5	2.2	11.5	1.9
Picture Completion	9.4	2.1	10.6	2.1
Picture Arrangement	10.1	2.9	11.1	2.8
Block Design	11.1	2.5	12.1	2.7
Object Assembly	11.1	3.3	13.6	2.9
Digit Symbol	12.7	2.4	13.8	2.1
Verbal IQ	106.3	9.4	109.2	8.8
Performance IQ	109.5	16.0	120.6	14.0
Full Scale IQ	109.0	12.2	115.6	11.1

Table 4

ANCOVA Summary Table for Full Scale IQ

Source of Variation	SS	df	MS	F
Error	562.07	41	13.71	--
First Testing Score	4698.73	1	4698.73	342.75
Gender	55.49	1	55.49	4.05*

* $p < .06$

Table 5

ANCOVA Summary Table for Verbal IQ

Source of Variation	SS	df	MS	F
Error	552.98	41	13.49	--
First Testing Score	2926.07	1	2926.07	216.95
Gender	125.48	1	125.48	9.30*

* $p < .01$

Table 6

ANCOVA Summary Table for Performance IQ

Source of Variation	SS	df	MS	F
Error	1944.89	41	47.44	--
First Testing Score	6381.22	1	6381.22	134.52
Gender	.59	1	.59	.01

Chapter 4

Discussion

The purpose of the present study was to assess the magnitude of practice effects on the reliability of the Wechsler Adult Intelligence Scale - Revised (WAIS-R) and to establish the stability of one of the normed age levels. Specifically, this study utilized 18 and 19 year olds, one of the age groups Wechsler neglected publishing normative data for in the WAIS-R manual. Indeed, Wechsler suggested further study be done on the WAIS-R for groups of people who were of differing age and ability levels. Through previous research, as noted in Chapter 1, the reliability of the WAIS-R may be suspect at varying age and ability levels requiring further research.

The findings from the present study indicate the WAIS-R is a highly reliable instrument for use in the 18 and 19 age groups. Through statistical analysis, the retest stability for Verbal, Performance, and Full Scale IQs for the combined sample was found to be comparable to the data published by Wechsler in the WAIS-R manual (1981) for the 25 to 34 and 45 to 54 year old age groups. Although the IQ score coefficients were similar to the coefficients obtained by Wechsler in his normative study, several individual subtest correlations in the present study differed from those obtained by Wechsler.

If a larger sample size corrects for most extreme scores, then the smaller sample would evidence the greater variability in scores and, consequently, a lower stability coefficient. The reasons for the contrary results are unclear; however, some tentative hypotheses may be formulated. The men who did volunteer may have constituted a self-

selecting group, one that may have been more similar in intellectual functioning than were the females in the sample.

This hypothesis gains credence when the results of the study are perused more closely. The retest stability coefficients for the men's Verbal, Performance, and Full Scale IQs (see Table 1) were slightly higher than the corresponding women's. This occurred despite there being markedly fewer men relative to women. Contrary to Wechsler's findings, females evidenced no increase in mean subtest scores on two subtests. These results may suggest that the present study's sample of men was quite similar in intellectual development, more so than the larger (and possibly more representative) sample of women.

As hypothesized, males evidenced significantly greater gains on Verbal and Full Scale IQs while Performance IQ showed no gender effect. It should be noted, however, that although statistical analysis yielded significance for retest gains on men's Full Scale IQs, the larger male Verbal IQ gain may have inflated the Full Scale IQ's F value and thus may render this statistically significant difference less meaningful.

To summarize, results of this study clearly indicated that the WAIS-R is a very stable and reliable instrument in the 18 and 19 age group. Two points of interest in the present study's findings were that the WAIS-R was very stable for both men and women. As predicted, the men had significantly greater gains from test to retest on Verbal and Full Scale IQ while no gender difference was apparent on Performance IQ.

Some element of caution should be used when reviewing this study. Additional research, including a more representative sample, is needed to further investigate the stability of the WAIS-R in the 18 and 19 age group. Also, specific attention might also be paid to the gender differences noted in the present study.

References

Ankney, C.D. (1992). Sex differences in relative brain size: The mismeasurement of woman, too? Intelligence, 16, 329-336.

Atkinson, L., Bowman, T.G., Dickens, S., Blackwell, J., Vasarhelyi, J., Szep, P., Dunleavy, B., MacIntyre, R. & Bury, A. (1990). Stability of Wechsler Adult Intelligence Scale-Revised factor scores across time. Journal of Consulting and Clinical Psychology, 2, 447-450.

Bayley, N. (1955). On the growth of intelligence. American Psychologist, 10, 805-818.

Boone, D.E. (1992). Reliability of the WAIS-R with psychiatric inpatients. Journal of Clinical Psychology, 48, 72-76.

Bootzin, R. R., Acocella, J. R., & Alloy, L. B. (1993). Abnormal psychology: Current perspectives (6th ed.). New York: McGraw-Hill Inc.

Dai, X., Ryan, J.J., Paolo, A.M., & Harrington, R.G. (1991). Sex differences on the Wechsler Adult Intelligence Scale-Revised for China. Journal of Consulting and Clinical Psychology, 3, 282-284.

Ilai, D., & Willerman, L. (1989). Sex differences in WAIS-R item performance. Intelligence, 13, 225-234.

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

Kaufman, A.S., McLean, J.E. & Reynolds, C.R. (1991). Analysis of WAIS-R factor patterns by sex and race. Journal of Clinical Psychology, 47, 548-556.

Klonoff, H. (1972). IQ constancy and age. Perceptual and Motor Skills, 35, 527-534.

Kowalski, J.M., & Rossini, E.D. (1990). Reliability of the WAIS-R factors. Psychological Reports, 66, 111-114.

Lubin, B., Larsen, R. M., & Matarazzo, J. D. (1984). Patterns of psychological test usage in the United States: 1935-1982. American Psychologist, 39, 451-454.

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

Matarazzo, J.D., & Herman, D.O. (1984). Base rate data for the WAIS-R: Test-retest stability and VIQ-PIQ differences. Journal of Clinical Neuropsychology, 6, 351-366.

Moore, A.D., Stambrook, M., Hawryluk, G.A., Peters, L.C., Gill, D.D., & Hymans, M.M. (1990). Test-retest stability of the Wechsler Adult Intelligence Scale-Revised in the assessment of head-injured patients. Journal of Consulting and Clinical Psychology, 2, 98-100.

Paolo, A.M. & Ryan, J.J. (1993). Test-retest stability of the Satz-Mogel WAIS-R Short Form in a sample of normal persons 75-87 years of age. Archives of Clinical Neuropsychology, 8, 397-404.

Piotrowski, C., & Keller, J. (1989). Psychological testing in outpatient mental health facilities: A national study. Professional Psychology: Research and Practice, 20, 423-425.

Quereshi, M.Y. & Ostrowski, M.J. (1985). The comparability of three Wechsler Adult Intelligence Scales in a college sample. Journal of Clinical Psychology, 41, 397-407.

Reynolds, C.R., Chastain, R.L., Kaufman, A.S. & McLean, J.E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardization sample as a function of the stratification variables. Journal of School Psychology, 25, 323-342.

Ryan, J.J., Georgemiller, R.J., Geisser, M.E., & Randall, D.M. (1985). Test-retest stability of the WAIS-R in a clinical sample. Journal of Clinical Psychology, 41, 552-556.

Schuerger, J.M., & Witt, A.C. (1989). The temporal stability of individually tested intelligence. Journal of Clinical Psychology, 45, 294-302.

Slate, J.R., Frost, J., & Cross, B. (1991). WAIS-R stability for college students with learning disabilities. Learning Disability Quarterly, 14, 2-6.

Snow, W.G., Tierney, M.C., Zorzitto, M.L., Fisher, R.H. & Reid, D.W. (1989). WAIS-R test-retest reliability in a normal elderly sample. Journal of Clinical and Experimental Neuropsychology, 11, 423-428.

Thompson, A.P., & Molly, K. (1993). The stability of WAIS-R IQ for 16-year-old students retested after 3 and 8 months. Journal of Clinical Psychology, 49, 891-898.

Watkins, C.E., & Campbell, V.L. (1992). The test-retest reliability and stability of the WAIS-R in a sample of mentally retarded adults. Journal of Intellectual Disability Research, 36, 265-268.

Wechsler, D. (1950). Intellectual development and psychological maturity. Child Development, 21, 45-50.

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

Wechsler, D. (1981). WAIS-R manual. San Antonio: The Psychological Corporation.

Willerman, L., Schultz, R., Rutledge, J.N., & Bigler, E.D. (1992). Hemisphere size asymmetry predicts relative verbal and nonverbal intelligence differently in the sexes: An MRI study of structure-function relations. Intelligence, 16, 315-328.

Appendix A

Participation Consent Form

Participation Consent Letter

Please read this participation letter carefully. You are invited to participate in a study investigating the stability of the Wechsler Adult Intelligence Scale - Revised (WAIS-R). The selection criteria are: (1) to be an 18 to 19 year old college student and (2) not to have taken the WAIS-R at any time prior to participation in this study. The individual conducting this study is a clinical psychology major in pursuit of a Master's degree.

Your participation will require approximately 60 to 90 minutes for administration of the WAIS-R and a subsequent administration of the WAIS-R two to eight weeks following the initial testing. If the participant desires, he/she may learn his/her scores on the WAIS-R. Your scores will remain confidential. Anonymity will be preserved, and only group scores will be reported.

Participation in this study is completely voluntary. If you wish to terminate your participation, you may do so at any point in the study. There is no risk or discomfort involved in completing the study. The benefits from your participation are the experience of being involved in a thesis study and the introduction to the administrative procedures of the WAIS-R. In addition, extra credit will be given by the course instructor upon completion of the test-retest procedure.

If you would like to volunteer, please sign up and include your telephone number. The author of this thesis will call you to schedule the second testing time. Please retain this consent form and bring it to your appointment. The examiner will sign the consent form thus verifying your participation, you may then return the bottom portion to your instructor in order to receive proper credit.

If you have any additional questions or comments about this study, feel free to contact **Matt Schrader** at 343-3265 or **Heather Kirchhefer** at 343-8821.

THANK YOU FOR YOUR PARTICIPATION!!!

I, _____, have read the above information and wish to participate
(please print name)
in this study. I understand that my participation is voluntary and that I may withdraw at any time without prejudice after signing this form should I choose to discontinue participation in this study. I understand that all information will be reported in a group format in order to provide confidentiality.

_____ _____
(signature of participant) (date)

_____ _____
(signature of examiner) (date)

THIS PROJECT HAS BEEN REVIEWED BY THE EMPORIA STATE UNIVERSITY COMMITTEE FOR THE PROTECTION OF HUMAN PARTICIPANTS.

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

FROM: Emporia State University Graduate School

I, Matthew S. Schrader, hereby submit this thesis/report to Emporia State University as partial fulfillment of the requirements for an advanced degree. I agree 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.

Matthew S. Schrader
Signature of Author

December 13, 1994
Date

The Test-Retest Reliability of the WAIS-R for the 18 and 19 Year Age Group
Title of Thesis/Research Project

Signature of Graduate Office Staff Member

December 14, 1994
Date Received

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