AN INVESTIGATION OF CHANGE/STABILITY IN WAIS-R SCORES FOR 16 YEAR OLDS OVER AN 18 MONTH PERIOD

Karen L. Molly©
Lakehead University

Thesis Supervisor: Dr. Tony Thompson

Submitted as partial fulfilment of the Master's
Degree in Clinical Psychology
at Lakehead University

September, 1992

National Library of Canada

Acquisitions and Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada

Direction des acquisitions et des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

ISBN 0-315-78943-3

Canada

Acknowledgements

Very special thanks to my Supervisor, Dr. Anthony Thompson, for his guidance and dedication over the course of this research. I would also like to thank Chris Rickards for her assistance in testing. Finally, the author acknowledges valuable comments by Dr. E. Bauman and Mr. James Cyr who carefully reviewed the manuscript.

Table of Contents

List of Appendices

List of Tables

Abstract

This study used a test-retest paradigm to investigate change/stability in Wechsler Adult Intelligence Scale-Revised scores of 16 and 17 year olds. Both short (3 month) and long (18 month) retest intervals were investigated with 26 subjects in each group. The results of this study were quite consistent in showing that WAIS-R retest gains for young 16 year olds were greater over an 18 month retest period than over a 3 month retest period. This general finding was true for males and females on mean Verbal IQ and Full Scale IQ. Mean Performance IQ also showed this differential retest effect but only for males. Verbal IQ gain for the long term retest group seems largely the result of the Vocabulary and Comprehension subtests. Some of the results varied by gender. Although absolute retest gains were different for the long versus short term retest groups, test-retest reliability was high for both groups. Absolute IQ gains for Performance and Full Scale were significantly, but not highly correlated with initial IQ score. Clinical and theoretical implications of these results were discussed. Further research needs to address change/stability of measured IQ for 16 and 17 year olds no longer attending school, and explore the relationship between gender and IQ change for this age group.

Introduction

The assessment of intelligence is a topic of much interest due to its scientific relevance and practical utility. Numerous theoretical, psychometric and clinical issues are associated with intellectual measurement. This study is concerned with changes in intellectual ability or, to view the issue from a related perspective, the stability of measured IQ. In particular, this study investigates the issue of change/stability in Wechsler Adult Intelligence-Scale-Revised (Wechsler, 1981; WAIS-R) scores of 16 year olds over an 18 month period. The literature relevant to this topic is first reviewed.

Growth and Decline of Intelligence

Studies investigating the nature of the complex relationship between aging and changes in intellectual abilities present mixed results. Based on some of the earlier research into the age of intellectual maturity, an adult level of intelligence was presumed to be attained by 16 or 17 years of age (Terman, 1917, 1937). Consequently, some tests provided one set of adult norms for all ages above 16 years since variability between age groups was thought to be minimal (Terman, 1917; Bloom, 1964; Thorndike, Hagen & Sattler, 1986;

Sattler, 1988). However, considerable evidence exists which suggests that for many individuals growth in intellectual abilities continues well into early adulthood (Freeman & Flory, 1937; Bayley, 1949, 1957; Kangas & Bradway, 1971; Matarazzo, Wiens, Matarazzo & Manaugh, 1973; Minton & Schneider, 1980; Kaufman, 1990). In general, cross-sectional research of adult intelligence reveals a peak in measured IQ between the ages of 18 and 25 years. Progressively lower group scores follow up to the age of 50, at which time the decline becomes even more pronounced. Longitudinal research tends to indicate a different pattern of development. Typically, greater increases in measured IQ during young adulthood are documented, proceeded by only a slight decline in intelligence which begins much later in middle adulthood (Kaufman, 1990; Roediger, Capaldi, Paris & Polivy, 1991).

Not all intellectual abilities appear to grow or decline over time. Horn and Cattell (1966) were the first to distinguish between fluid and crystallized intelligence. Fluid intelligence is the general ability to perceive, encode and reason about information. It derives from biological and genetic factors and is less influenced by education and experience. Such abilities appear to increase with neurological maturation during childhood and adolescence and show decline throughout later adulthood. Crystallized intelligence involves the ability to understand relationships and solve problems and is based more on education and experience. These abilities appear to increase steadily across the life span and are less likely to show a decline with age (Horn & Cattell, 1966; Kaufman, 1990; Roediger et al, 1991). In general, sub-tests on the WAIS-R Verbal Scale are thought to measure educational-related abilities associated with Horn and Cattell's crystallized intelligence. Tests on the Performance Scale assess problem-solving abilities characteristic of Horn and Cattell's fluid intelligence (Kaufman, 1990).

The growth and decline of intellectual abilities is a function of the complex interplay of genetic and environmental influences. Research addressing the extent to which genetic and environmental influences account for variation in measured IQ between individuals, families, generations and cultures is inconclusive and certainly controversial. General intelligence is currently estimated to have an average heritability index of 52% (Kaufman, 1990) although estimates have been as high as 88% over the years (Bloom, 1964). The current index suggests that approximately half the observed variance in IQ scores in the general population is attributable to the influence of genetic factors (Sattler, 1988).

Although heredity sets limits on general intellectual functioning, the interaction of environmental factors determines if such potential is realized (Sattler, 1988). Variations in the environment tend to have the greatest quantitative effect on an ability at its most rapid period of development or change (Bloom, 1964). Specific aspects of the environment which have been reported as potentially significant factors in influencing general intellectual functioning include: neonatal and general birth processes; birth order; level of education; and home environmental variables such as socioeconomic status, cognitive stimulation, achievement orientation and motivation. (Sattler, 1988; Kaufman, 1990). Taken in isolation, any one factor does not account for large proportions of IQ variance in the general population. Rather, the environment presents a complex interaction of numerous influences which, when taken together, account for a large percentage of IQ variance (Bloom, 1964; Sattler, 1988; Kaufman, 1990).

There is considerable individual variability within the group patterns of change typically associated with various ages. Genetically-based developmental trends and environmental factors can result in significant shifts in measured IQ for some individuals during certain stages of growth and over a wide range of intellectual abilities (Freeman & Flory, 1937; Bayley, 1949, 1957; Bloom, 1964; Sattler, 1988). Subsequently, two persons of the same measured ability at a given time may differ markedly in that ability at a future time if growth rates (e.g. a continuous rate or one which occurs in spurts and pauses) and environmental influences are significantly different. Variability in measured IQ between individuals decreases as intellectual maturity is reached (Bloom, 1964; Thorndike et al, 1986; Sattler, 1988).

The Stability of Measured IQ

Research into the stability of measured IQ addresses the relative position of individuals within a group overtime. Absolute difference between groups overtime is the focus of the growth and decline research. Stability refers to correlations obtained for the same group of individuals measured at various times. Research investigating the stability of measured IQ indicates that intellectual ability relative to similar aged peers remains fairly constant after reaching school age (Terman, 1917; 1937; Bayley, 1949; McCall, Appelbaum & Hogarty, 1973; Sattler, 1988; Schuerger & Witt, 1989; Roediger, et al, 1991). Test-retest reliability increases as age of initial testing increases. Specifically, mean correlations between childhood and adult IQ increase as childhood IQ is measured closer to school age. For example, mean correlations between IQ scores in childhood and IQ at the age of 18 rise significantly between the ages of 3 and 6, with correlations reported as high as .80 and above after age 6. Reliability typically tends to decrease as the interval between testing increases (Bloom, 1964; McCall et al, 1973; Schuerger & Witt, 1989).

The poor predictive validity of scores on intelligence tests given to very young children is a reflection of the different kinds of items used on tests at various age levels. For example, infant scales are primarily of a perceptual-motor nature. Such tasks as stringing beads or identifying bodily parts may not be related to tasks included on intelligence tests for older children and adults such as vocabulary and reasoning (Roediger et al, 1991). Preschool intelligence tests, however, contain more items reflecting cognitive ability and subsequently have greater predictive power. For this reason, the constancy of the IQ score is influenced considerably by the age of the child at initial testing. The older the child, the greater the constancy of score (Sattler, 1988).

Test-retest stability of Wechsler's intelligence scales, particularly the WAIS-R, is relevant to this thesis research. In general, the Wechsler scales have good reliability as reflected by high retest correlations over short and longer testing intervals (Kangas & Bradway, 1971; Matarazzo et al, 1973; Brown & May, 1979; Catron & Thompson, 1979; Wagner & Caldwell, 1979; Matarazzo, Carmody and Jacops, 1980; Wechsler, 1974, Wechsler, 1981; Matarazzo & Herman, 1984; Schuerger & Witt, 1989; Kaufman, 1990). Previous research findings for the Wechsler Adult Intelligence Scale (WAIS) are generally applicable to the WAIS-R (Kaufman, 1990). Matarazzo and colleagues (1980) reviewed 11 studies with retest intervals on the WAIS ranging from one week to 13 years. Subjects ranged in age from 19 to 70 years. Verbal, Performance and Full Scale IQs had median stability coefficients of .89, .85 and .90, respectively. High stability scores were found as frequently in studies which utilized longer retest intervals as studies using shorter intervals. Furthermore, retest stability was as high for one age level as another. Wechsler (1981) presents test-retest data on the WAIS-R for 119 adults ranging in age from 25-34 and 48-54. Test-retest intervals were from 2 to 7 weeks. Reliabilities averaged .95 for Verbal and

Full Scale IQ and .90 for Performance IQ.

In summary, several points emerge from the research on intellectual growth/decline and the stability of measured IQ over time. There is a general group trend of continued intellectual growth into later adolescence and early adulthood. There is a differential pattern of growth and decline in group intellectual abilities across the age span associated with Horn and Cattell's fluid and crystallized intelligence. There is considerable individual variability within the patterns of group change associated with various ages. Such trends are a function of genetic and environmental influences. Group trends in measured intelligence indicate stability of scores over time. Specifically, correlations between childhood and adult IQ increase as childhood IQ is measured closer to school age. Wechsler scales, and particularly the WAIS-R have good reliability as reflected in high test-retest correlations.

Methodological Issues in the Assessment of Intelligence Over Time

Cross-Sectional and Longitudinal Designs

Much of the research into issues of change /

stability in intellectual abilities across the life span is cross-sectional in nature. Different groups of individuals of specific ages are tested at the same point in time. However, these groups are not samples of the same population due to differential effects of factors such as historical events, the impact of mass media, health and medical care, child rearing techniques and educational attainment (Roediger et al, 1991; Kaufman, 1990). These time-related influences are referred to as cohort or generational effects and are thought to significantly contaminate the comparison of different age groups using the cross-sectional approach. Cross-sectional research by Flynn (1984; 1987) indicates significant gains in measured IQ across generations in 14 nations. These gains differ substantially from country to country and are considered to reflect cultural/environmental influences or the differential effects of cohort on measured IQ (Flynn, 1984; 1987; Kaufman, 1990). An additional limitation with cross-sectional research is that trends found in groups may not always apply to individuals (Freeman & Flory, 1937; Kaufman, 1990).

The longitudinal method of research better addresses the issue of change/stability in measured intelligence. Variance attributable to the differential influence of cohort is held constant as the same individual or group of individuals is tested repeatedly over time. However, a problem inherent to this design is that when the same individual is repeatedly tested there is a practice effect that differentially influences scores on both the Verbal and Performance scales (Kaufman, 1991). Furthermore, the current practice of computing IQ scores using the deviation from comparison age group complicates the issue of the measurement ofintellectual growth. Change in measured IQ across testings may indicate "real" growth in intellectual ability or reflect change in the age-relevant standard.

Practice Effects

Longitudinal research investigating change/stability in measured IQ has inherent to its design the problem that an increase in score across testings may reflect experience at taking intelligence tests versus "real" gains in intellectual ability. Research addressing the effect of retaking an intelligence test indicates that there are significant gains in IQ scores upon retesting (Kangas & Bradway, 1971; Matarazzo et al, 1973; Catron, 1978; Catron & Thompson, 1979; Matarazzo et al, 1980; Matarazzo & Herman, 1980; Shatz, 1981; Wechsler, 1981; Matarazzo & Herman, 1984; Schuerger and Witt, 1989; Kaufman, 1990).

Such retest gains are classically referred to as practice effects. Literature on the effects of practice on measured IQ is relevant to the current research topic.

A study by Catron and Thompson (1979) explored the relationship of WAIS retest gains to the interval between testings in college undergraduates. Subjects were tested at one, two, three or four month intervals. There were significant gains in retest scores across all intervals, with the exception of Verbal IQ gain at four months which was not significant. Average gain in Performance IQ exceeded gains in Verbal IQ across all testings. Gains decreased as the length of the interval increased. Average gain scores on Verbal IQ over each of four successive one month intervals were 4.74, 1.79, 2.27 and .85 respectively. Similar gain scores on Performance IQ were 11.37, 9.79, 7.74 and 8.00. Full Scale IQ gains across testings were 8.00, 5.68, 5.42 and 4.21.

Change in measured IQ across testings on the WAIS and WAIS-R scales has been thoroughly reviewed and investigated by Matarazzo and colleagues (1973; 1980) and Matarazzo and Herman (1984a). These authors conclude that there is a very profound retest effect on measured IQ. This effect is particularly pronounced on the Performance scale and over shorter retest intervals. Among separate WAIS-R subtests the largest practice effect is for the Object Assembly and Picture Arrangement subtests. The smallest retest gain is for Vocabulary and Information. The effects of practice occur across individuals and groups. Significant decreases in individual IQ score on retesting are rare. Finally, there was no relationship between initial IQ score and size of gain on retesting.

Using data from Wechsler's (1981) study, Matarazzo and Herman (1984) provided the following mean group retest gain scores that could be expected when the initial test and subsequent retest are with the WAIS-R: Verbal IQ gain scores of 3 points; Performance IQ gains of 8 points and Full Scale IQ gain scores of 6 points. These scores are most generalizable to adults aged 25-34 and retest intervals of one through six months. Based on Matarazzo and Herman's (1984) data the following gains in IQ score would be necessary to infer significant improvement in intellectual ability in individuals across testings. At the 5% criterion level, an increase on the Verbal IQ scale of 12 points is required with gains in Performance IQ approximating 23 points for statistical significance. Retest gains on Full Scale IQ should be 15 points or one standard deviation to infer a significant improvement in ability. At the 10% criterion level, Verbal IQ gains should exceed 10 points, Performance IQ 19 points and Full Scale IQ 13 points. Although loss in score on retesting is rare, Matarazzo and Herman (1984a) provide the following minimum decrease in IQ score (at the 5% criterion level) necessary to infer a statistically significant loss in function: Verbal IQ 5 points; Performance and Full Scale IQ 4 points. Although a particular sub-test is not interpreted in isolation, a change of 3-5 points on retesting may be interpreted as clinically significant(Matarazzo et al, 1980).

The typically large retest effect on Performance IQ is likely a result of the examinee developing problem-solving strategies which can be applied to the same or similar problems on future testings. Also contributing to the larger retest effect on the performance scale is the importance of speed in calculating scores (Kaufman, 1990).

Further research is necessary to understand how the effects of practice relate to individuals of different populations, ages, educational levels and initial IQ score. For example, a study by Shatz (1981) provides evidence suggesting that the effects of practice are considerably smaller in elderly individuals, especially those with organic brain damage. A study by Bauman (1991) suggests that IQ scores of children with learning difficulties may not show the same retest effect found in a normal population. Group means for Verbal and Full Scale IQ decreased over testings. Many children in Bauman's research experienced statistically significant declines in Verbal and Full Scale IQ scores. There was a small but significant mean gain on Performance IQ.

Extraneous Variables

There are a number of extraneous variables which can further influence the measurement of intelligence over time and the change/stability of IQ scores. Situational factors and scorer error are frequently cited as potentially significant influences on measured IQ (Sattler, 1988; Kaufman, 1990). Situational variables encompass factors related to both the examiner and examinee. One important source of situational bias is the degree of rapport established between the examiner and examinee. For example, research indicates that discouragement during an examination lowers the scores obtained by children (Sattler, 1988). The effects of encouragement are less likely to result in significant examiner effects (Sattler, 1988). Other factors which can affect examinee performance include examiner personality (Moriarty, 1966); examinee physical health, fatigue, anxiety, level of motivation and self-confidence

(Sattler, 1988; Kaufman, 1990). A few studies have evaluated the role of examiner sex finding no systematic influence on IQ scores(Sattler, 1988). While situational factors can influence test scores, they result in smaller changes in IQ scores then scorer errors.

Research addressing the prevalence and impact of scorer error on measured IQ indicates that it is very common for examiners to make errors in scoring (Brannigan, 1975; Wechsler, 1981; Ryan, Prifitera & Powers, 1983; Slatter, 1988; Kaufman, 1990; Slate & Jones, (1990). The experience level of the examiner is unrelated to scoring accuracy. Both experienced and in-experienced examiners make similar errors. In fact, research suggests experienced examiners make more errors than those with less experience (Brannigan, 1975; Ryan et al, 1983; Sattler, 1988). Incorrectly crediting test items is a large source of scorer error variance. This is particularly true on Verbal sub-tests which require a large degree of judgment by the examiner(Wechsler, 1981; Kaufman, 1990). More errors are made on the Vocabulary subtest followed by the Comprehension and then the Similarities subtests. Such errors when made tend to be biased towards leniency, significantly inflating IQ scores (Wechsler, 1981; Sattler, 1988; Kaufman, 1990; Slate & Jones, 1990).

Warren and Brown (1973) report errors in FSIQ as great as 5 points in 47% of protocols given by a sample of graduate students. In a study by Slate and Jones (1990), 22 Master's level students in Clinical Psychology scored 7 WAIS-R protocols. Students made on average 7.95 errors per protocol. Corrected protocols indicated students overestimated 56% of FSIQ's, ranging from 1-10 IQ points, and underestimated 16%, ranging from 1-2 points. Subtests having the greatest number of errors (over-estimates) were Vocabulary (M=2.68); Comprehension (M=1.78) and Similarities (M=0.97).

Computational error in scoring is another major source of scorer error and include miscalculation in the addition of raw scores and in the conversion of raw scores to scaled scores. In a study by Ryan et al (1983), 19 psychologists and 20 graduate students produced summary scores for two vocational counselling clients which varied as much as 4 to 18 IQ points. In summary, situational variables and scorer errors can result in considerable variability in measured IQ, significantly decreasing the reliability and validity of obtained scores on the WAIS-R. The standards of measurement provided in test manuals are based on internal consistency and do not adequately take subjective variables into consideration. During standardization, the test protocols are scored and rescored by statistical clerks to control for accuracy (Kaufman, 1990; Slate & Jones, 1990). Reliability studies need to pay special attention to establishing comfortable levels of rapport and accuracy/consistency in scoring to minimize these influences.

Intellectual Assessment of 16 and 17 Year Olds

Research specifically addressing the issue of change/stability of measured IQ in 16 and 17 year olds is limited. Almost five decades ago, Knezevich (1946) investigated the issue with one hundred and thirteen rural secondary school students. The students were initially tested with the Henmon-Nelson Test of Mental Ability (Forms A,B,C) in their sophomore year of high school, and then again when they were seniors with parallel forms of the test. The mean age in the sophomore year was 182.8 months (15 years, 3 months) and 206.76 months (17 years, 3 months) when retested in the senior year. The mean IQ in the sophomore year was 104 (S.D=10.96). The mean IQ in the senior year was 106.4 (S.D=8.05). The IQ score of 8 individuals remained the same over testing. Sixty-three individuals showed a gain in IQ; 61% of the gains were greater than 5 points. Forty-three cases showed a loss in IQ score on retesting, with 42% declining more than 5 points. The Full Scale test-retest reliability coefficient was .70. Based on a comparison of chronological age and mental age, the authors concluded that mental growth did not stop during this period for many individuals.

A classic cross-sectional study by Bayley (1957) addressed the issue of growth in intelligence between the ages of 16 and 21 as measured by the Wechsler-Bellevue Scale. The 33 subjects in this study had been part of the Berkeley Growth Study, a long-term developmental investigation of mental, motor and physical development (Jones & Bayley, 1941). As a result, Bayley's subjects had received multiple assessments (approximately 36) since infancy and up to 12 years of age. In the investigation of teens, subjects were administered the Wechsler-Bellevue Adult Intelligence Scale (Form 1) at 16, 18, and 21 years of age. There were significant gains in IQ score at all levels of intelligence and across all levels of education. There were no gender differences. Thirty of the thirty-three individuals tested had gains in IQ score ranging from 1 - 20 points. The greatest magnitude of gain was in the 16 -18 age interval. Mean Full Scale IQ scores increased five points between 16 - 18 years of age and a further two points from 18 - 21. These five year gains (16 to 21 years) are reportedly significant at the .001 level of confidence (Bayley, 1957).

Test-retest correlations were also high in Bayley's study. For Full Scale IQ, the correlation between the ages of 16 and 18 years of age was .96. Retest correlations on the Verbal and Performance Scales were .90 and .86 respectively (Bayley, 1957). Although there was no clear relationship between initial IQ level and gain in intellectual growth over time, Bayley noted that the scatter of scores indicate that those individuals with originally low or average IQs appeared to gain steadily over the five year retesting interval (Bayley, 1957).

Current Investigation & Hypotheses

The present research used a test-retest paradigm to investigate change/stability in IQ scores for 16 and 17 year olds. Both short (3 month) and long (18 month) retest intervals were investigated. The issue of IQ change/stability was not confounded in this research by different tests or test norms. There were two hypotheses. First, it was predicted that mean IQ scores would increase in both retest groups, but that the long term retest group would gain more than the short retest group. Second, it was hypothesized that test-retest reliabilities would be high in both groups.

Method

Subjects: Sixty-eight subjects were recruited from four local public high schools over a four month period. Because the age level considered appropriate for the WAIS-R is 16 years, the age criterion for acceptable subjects was set at a minimum of 15 years, 11 months and 15 days. One subject 15 years, 11 months, 11 days was included. Ten subjects under the minimum age criterion (ranging in age from 15 years, 10 months, 7 days to 15 years, 10 months, 27 days) were given the first administration of the WAIS-R due to a misunderstanding by the researcher about the minimal age criterion. These ten subjects were subsequently dropped from the study. Of the remaining 58 subjects (25 males, 33 females), complete data (test and retest) were obtained for 52 subjects (22 males, 30 females) which made up the final sample for this study. These subjects ranged in age from 15 years, 11 months, 11 days to 16 years, 3 months, 9 days at the time of initial testing.

Measure: The Wechsler Adult Intelligence Scale Revised (WAIS-R) was administered to all subjects on two separate occasions. The WAIS-R was administered as it is generally viewed as the standard for the assessment of adult intelligence and is subsequently the most commonly used test of intelligence (Harrison, Kaufman, Hickman & Kaufman, 1988; Archer, Marwish, Imhof & Piotrowski, 1991).

Procedure: Ethical approval for this project was granted by the Lakehead University Ethics Advisory Committee (see Appendix A). Permission was also obtained from the Lakehead Board of Education to approach students in local high schools (see Appendix A). Subjects were randomly assigned to the short or long retest condition. As far as possible, subjects were assigned alternately to examiners although this was constrained by examiner availability. Although random assignment of subjects to condition was not stratified by sex, it turned out that the same number of males and females were in each group. The target short retest interval was three months and the target long retest interval was 18 months. Eighteen months was chosen as subjects had to be retested prior to their 18 birthday so that the same test norms for 16 and 17 year olds could be used. Furthermore, 18 months marked the end of the academic year and optomized availability of participants. A Consent to Participate Form was signed by the participant's parent or guardian. (see Appendix B). All subjects were tested and retested by one of two female M.A. candidates. Both examiners had completed a graduate psychometric assessment course which included instruction on the WAIS-R and a competency test. Test administration was further reviewed and practised with the thesis supervisor. The same examiner administered and scored the first and second WAIS-R for all subjects except three in the long retest group. Each examiner tested roughly an equivalent number of subjects in the short (15 versus 11) and long (16 versus 10) retest conditions. Special attention was given to establishing good rapport with each subject and to scoring the test protocols accurately and consistently across testings. Upon retesting, each subject was contacted by telephone and provided with feedback based on their performance on the initial test. The general feedback protocol is provided in Appendix C. The ten subjects dropped from the study in its early stage because of their inappropriate age were also provided with feedback (See Appendix D).

Results

Retest Summary Statistics:

Summary data for the short and long term retest procedure are summarized in Table 1. There were 11 males and 15 females in each group at Time 2. The mean age of subjects in the short and long term retest group at Time 1 was 16.12 years and 16.06 years, respectively. The age range of the total sample at

Table 1

Sample Characteristics and Retest-Time Data for Short and Long Term Retest Groups

	Short Retest Group		Long Retest Group	
	Time 1	Time 2	Time 1	Time 2
<u>Subjects</u>				
Total(N)	26	26	26	26
Male	11	11	11	11
Female	15	15	15	15
<u>Mean Age</u>				
Total(years)	16.12	16.37	16.06	17.58
Male	16.09	16.35	16.07	17.58
Female	16.13	16.39	16.06	17.57
<u>Age Range</u>				
Total (Y-M-D)	15-11-17 to 16-03-09	16-02-10 to 16-06-15	15-11-11 to 16-02-09	17-05-16 to 17-08-16
Male	15-11-17 to 16-03-04	16-02-29 to 16-06-15	15-11-18 to 16-02-09	17-05-24 to 17-08-16
Female	15-11-18 to 16-03-09	16-02-10 to 16-06-11	15-11-11 to 16-02-04	17-05-16 to 17-08-16
<u>Retest Time</u>				
Mean(months)	3.04		17.77	
Range(M-D)	2-18 to 3-18		17-13 to 18-12	

Time 1 was 15 years, 11 months, 11 days to 16 years, 3 months, 9 days. The mean retest interval for the short term retest group was 3.04 months, with a range of 2 months, 18 days to 3 months 18 days. The mean retest interval for the long term retest group was 17.77 months, with a range of 17 months, 13 days to 18 months 17 days.

Verbal, Performance and Full Scale IQ

The means and standard deviations for Verbal, Performance and Full Scale IQ are depicted in Table 2 broken down by time of testing, retest condition and gender. Independent t-tests were performed to compare the short and long term retest groups on Verbal, Performance and Full Scale IQ at Time 1. There were no significant differences in IQ found for males and females and for the gender-combined samples. These results indicate that the re-test groups were initially equivalent in terms of measured IQ.

Gender differences were also investigated for Verbal, Performance and Full Scale IQ within each re-test group at Time 1. Independent t-tests revealed males scored significantly higher than females on Verbal IQ in the short retest group ($t(24)=2.46$, $p<.02$). The long retest group was in the same direction, however fell just short of statistical

Table 2

Means and Standard Deviations for Verbal, Performance and Full Scale IQ by Time of Testing, Retest Condition and Gender.

		VIQ		PIQ		FSIQ	
	N	M	SD	M	SD	M	SD
Time 1							
Short Retest							
Males	11	100.73	9.59	102.09	11.18	101.00	9.58
Females	15	91.47	9.30	99.80	8.56	94.53	8.28
Total M/F	26	95.38	10.34	100.77	9.61	97.27	9.26
Long Retest							
Males	11	104.91	11.16	108.55	13.15	106.55	12.39
Females	15	96.00	10.91	103.67	12.11	98.93	10.82
Total M/F	26	99.77	11.69	105.73	12.54	102.15	11.90
Time 2							
Short Retest							
Males	11	103.73	8.75	111.36	10.94	107.27	9.89
Females	15	95.53	9.95	111.60	11.24	102.20	9.34
Total M/F	26	99.00	10.16	111.50	10.89	104.35	9.72
Long Retest							
Males	11	113.09	12.67	124.09	12.82	119.73	13.61
Females	15	102.67	12.78	113.33	14.24	107.40	13.43
Total M/F	26	107.08	13.54	117.88	14.44	112.62	14.62

significance at the .05 level of probability ($t(24)=2.03$, $p<.06$.)

To determine whether there were differential effects related to short and long term retesting, a separate Analysis of Variance was conducted for Verbal, Performance and Full Scale IQ at Time 2. There were two factors for each ANOVA; retest condition (short, long) and gender (male, female). Also, Verbal, Performance and Full Scale IQ at Time 1 served respectively as the covariate. For Verbal IQ, the only significant effect was for retest condition ($F(1,47)=8.16$, $p<.01$). This result indicates that males and females in the long term retest group increased significantly more on Verbal IQ than the short retest group, controlling for Verbal IQ at time of initial testing.

For Performance IQ, the only significant effect was the condition by gender interaction. Males in the long term retest condition increased on Performance IQ significantly more relative to males in the short term retest condition at Time 2 ($F(1,47)=4.24$, $p<.05$), controlling for Performance IQ at time of initial testing. Females showed no short versus long retest differences.

For Full Scale IQ the main effect for condition was significant ($F(1,47)=6.46$, $p<.01$) as was the condition by gender interaction effect ($F(1,47)=6.61$, p<.01). Males and females in the long term retest group increased significantly more than males and females in the short term retest group. Furthermore, this effect was stronger for males versus females.

Table 3 displays a distribution of individual change scores for Verbal, Performance and Full Scale IQ for the short and long term retest groups. Although these distributions are not broken down by gender to reveal the significant interaction effects of the ANOVA analyses, they do give an appreciation of the differential effect of long versus short retest intervals and are a key toward clinical relevance. The differential retest effect is also somewhat captured by the comparative mean gain scores
for the short and long term retest groups in Table 2. This was most evident for Verbal IQ on which the short term retest group gained 3.62 points and the long term retest group gained 7.31 points.

A correlational analysis indicated that initial IQ was correlated with retest gain for two of the three IQ indices. Gain in Performance IQ at Time 2 was significantly correlated with initial Performance IQ (r=.34, p<.05). Gain in Full Scale IQ at Time 2 was also significantly correlated with initial Full Scale IQ (r=.28, p<.05). These results indicate that although there is a relationship between initial

Table 3

Distribution of Individual Change Scores for Verbal, Performance and Full Scale IQ for Short and Long Term Retest Groups

	VIQ		PIQ		FSIQ	
Change Score	Short	Long	Short	Long	Short	Long
-8 to -5	1	0	0	1	0	0
-4 to -1	4	1	1	2	1	0
0 to 3	6	4	3	1	4	2
4 to 7	12	9	5	3	6	3
8 to 11	2	8	4	5	11	11
12 to 15	1	2	7	6	4	8
16 to 19	0	2	5	2	0	2
20 to 23	0	0	1	4	0	0
24 to 27	0	0	0	0	0	0
28 to 31	0	0	0	2	0	0

Performance and Full Scale IQ score and gain score, the correlation is not a strong one, accounting for less than 10% of total variance. Verbal IQ gain at Time 2 was not related to initial Verbal IQ score.

Test-Retest Reliabilities for Verbal, Performance and Full Scale IQ for the short and long term retest groups are summarized in Table 4. Using Fisher's r to z transformation, these reliabilities were not significantly different than those reported by Wechsler (1981).

Verbal and Performance Subtests

Scaled score means and standard deviations for the Verbal subtests by time of testing, retest condition and gender are shown in Table 5.

To determine whether there were differential effects related to short and long term retesting on the Verbal subtests, a Multivariate Analysis of Variance was conducted. There were two factors; retest condition (short, long) and gender (male, female). Full Scale IQ at Time 1 was used as a covariate in order to control for the effects of overall intelligence level. The only significant multivariate effect was for condition ($F(6,42)=3.04$, $P<.05$). This result indicates that males and females in the long term retest condition increased significantly more on the Verbal subtests than the short retest group,

Table 4

Test-Retest Reliabilities for Verbal, Performance and Full Scale IQ for Short and Long Term Retest Groups

	N	VIQ	PIQ	FSIQ
Short Retest	26	.91	.82	.92
Long Retest	26	.94	.79	.95
Wechsler[1]	71	.94	.89	.95

[1] Wechsler Manual (1981) - Re-test Reliabilities for 25-34 year olds, 2-7 week retest interval.

Table 5

Scaled Score Means and Standard Deviations for Verbal Subtests by Time of Testing, Retest Condition and Gender.

		Time 1				Time 2			
		Short		Long		Short		Long	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
Verbal Subtests									
Info.	M	7.64	1.69	8.55	1.75	7.73	2.15	9.45	2.01
	F	5.53	1.64	6.07	1.91	5.73	1.71	6.60	2.41
	T	6.42	1.94	7.12	2.20	6.58	2.12	7.81	2.64
DigitSp.	M	9.18	1.99	9.64	1.28	10.36	2.16	10.45	2.11
	F	9.33	2.44	9.67	2.09	9.53	2.53	10.80	2.62
	T	9.27	2.22	9.65	1.77	9.88	2.37	10.65	2.38
Vocab.	M	8.00	1.84	9.36	2.25	8.36	1.63	10.45	2.38
	F	6.73	1.28	7.60	1.96	7.33	1.40	8.40	2.16
	T	7.27	1.64	8.35	2.23	7.77	1.56	9.27	2.44
Arith.	M	8.64	2.01	9.27	2.00	8.91	1.70	10.36	3.11
	F	7.53	1.73	8.13	2.45	7.73	1.98	9.27	2.40
	T	8.00	1.90	8.62	2.30	8.23	1.92	9.73	2.72
Comp.	M	8.00	1.10	8.73	2.40	8.18	1.08	10.18	2.36
	F	6.73	1.87	7.47	1.68	7.47	1.60	8.80	1.74
	T	7.27	1.69	8.00	2.06	7.77	1.42	9.38	2.10
Simil.	M	9.64	2.70	9.18	2.56	10.27	2.53	10.64	1.91
	F	7.87	1.73	8.47	1.36	9.20	2.43	9.13	1.88
	T	8.62	2.30	8.77	1.95	9.65	2.48	9.77	2.01

M=Males; F=females; T=Males and Females

controlling for Full Scale IQ at time of initial testing. Univariate tests revealed that increases on Vocabulary ($F(1,47)=5.84$, $p<.05$) and Comprehension ($F(1,47)=9.36$, $p<.01$) subtests were largely responsible for the multivariate condition effect.

Scaled score means and standard deviations for the Performance subtests by time of testing, retest condition and gender are shown in Table 6.

To determine whether there were differential effects related to short and long term retesting on the Performance subtests, a Multivariate Analysis of Variance was again conducted. There were two factors, retest condition (short, long) and gender (male, female). Full Scale IQ at Time 1 was used as a covariate. The only significant effect was for gender ($F(6,42)=2.71$, $p<.05$). Overall males scored higher than females on the Performance subtests at Time 2, controlling for Full Scale IQ at time of initial testing. Univariate tests revealed that males scored higher than females on Picture Completion ($F(1,47)=3.60$, $p<.10$) and Digit Symbol ($F(1,47)=12.92$, $p<.01$).

Test-retest reliabilities for Verbal and Performance subtests for the short and long term retest groups are summarized in Table 7. The reliability coefficients are generally quite similar for the short

Table 6

Scaled Score Means and Standard Deviations for Performance Subtests by Time of Testing, Retest Condition and Gender.

		Time 1				Time 2			
		Short		Long		Short		Long	
		Mean	SD	Mean	SD	Mean	SD	Mean	SD
<u>Performance Subtests</u>									
PictC.	M	9.18	1.33	8.73	2.00	10.18	1.17	10.45	1.75
	F	8.13	2.33	8.07	2.15	8.87	2.39	8.73	1.94
	T	8.58	2.00	8.35	2.08	9.42	2.04	9.46	2.02
PictA.	M	9.00	1.73	10.18	2.40	10.36	2.62	11.45	2.16
	F	8.33	1.76	8.53	2.36	9.87	2.10	10.27	2.66
	T	8.62	1.75	9.23	2.47	10.08	2.30	10.77	2.49
BlockD.	M	11.18	2.09	12.09	2.07	12.09	3.70	14.00	2.19
	F	9.53	2.26	10.67	2.22	11.07	2.87	11.13	2.90
	T	10.23	2.30	11.27	2.24	11.50	3.22	12.35	2.95
ObjectA.	M	10.55	2.84	11.00	1.79	11.45	1.75	13.91	2.39
	F	10.33	2.64	10.60	3.25	12.27	2.84	12.13	2.92
	T	10.42	2.67	10.77	2.69	11.92	2.43	12.88	2.80
DigitSy.	M	8.82	2.09	10.09	2.74	9.82	3.12	11.36	2.20
	F	11.13	2.45	11.73	2.34	12.20	2.76	12.53	2.42
	T	10.15	2.54	11.04	2.60	11.19	3.10	12.04	2.36

M=Males; F=Females; T=Males and Females

Table 7

Test-Retest Reliabilities for Verbal and Performance Subtests for Short and Long Term Retest Groups

	Short	Long
Verbal Subtests		
Information	.88	.80
Digitspan	.80	.71
Vocabulary	.86	.90
Arithmetic	.73	.72
Comprehension	.73	.84
Similarities	.68	.76
Performance Subtests		
Picture Completion	.61	.58
Picture Arrangement	.53	.72
Block Design	.80	.88
Object Assembly	.68	.55
Digit Symbol	.84	.82

and long retest conditions. The largest coefficient discrepancy among the Verbal subtests was for Comprehension (short group r=.73; long group r=.84); Among the Performance subtests, the largest coefficient discrepancy was for Picture Arrangement (short group r=.53; long group r=.72). Using Fixer's r to z transformation, these reliabilities were not significantly different.

Discussion

The results of this study were quite consistent in showing that WAIS-R retest gains for young 16 year olds were greater over an 18 month retest period than over a 3 month retest period. This general finding was as hypothesized for males and females on mean Verbal IQ and Full Scale IQ. Mean Performance IQ showed this differential retest effect but only for males. It is important to acknowledge several considerations relevant to this main finding. First, the differential retest effect seems quite solid for a number of reasons. The same test was administered to the same individual on two separate occassions. Furthermore, the normed group for 16 and 17 year olds was the same. Subsequently, the differential retest effect is not a reflection of change in test design or the age-relevant standard.

IQ at time of initial testing was also used as a covariate to control for the effects of initial IQ on subsequent score. Various extraneous influences were minimized as the examiners paid special attention to establishing a good level of rapport with each subject and to scoring the test protocols accurately and consistently across testings.

The effects of practice on IQ score are considerable over shorter retest intervals, diminishing considerably over one year intervals. Mean short retest gains in this research on Verbal (3.62), Performance (10.78) and Full Scale IQ (7.08) are similar to the figures provided by Matarazzo and Herman (1984; VIQ=3; PIQ=8; FSIQ=6) using data from Wechsler's 1981 study. Average gains in the long term retest group on Verbal and Full Scale IQ were over and above those which could be attributable to the effects of practice in the short term retest group. Again, this is reason to be more confident about the long-term retest effect which would have arisen with minimal benefits from the effects of practice. Although individual change scores for the long term retest group (Table 3) are in most cases less than the clinically relevant criterion set by Matarazzo and Herman (1984), it can be argued that their standard is not appropriate for the long term retest group. The figures Matarazzo and Herman provide are for ages 25-34 and over short retest intervals (1-6 months), thereby including optimal practice effects. Change scores for the long term retest group would not be affected by practice to the same extent.

Verbal IQ gain for the long term retest group seemed largely the result of score increases on the Vocabulary and Comprehension subtests. Such Verbal subtests are thought to assess educationally-related abilities (Horn and Cattell's crystalized intelligence) and are reportedly less likely influenced by the effects of practice.

Gains in Performance and Full Scale IQ over retest interval were correlated with initial IQ score, but the relationship was very small accounting for only 10% of the total variance in gain scores. Verbal IQ gains at Time 2 were not related to initial IQ score. However, test-retest reliability were high for both groups as hypothesized. Thus, relative IQ position can be predicted well over 3 or 18 month intervals, but absolute gains over these time periods are not very predictable from initial IQ scores.

There were some unexpected retest effects that varied by gender. Males in the long term retest condition increased on Performance IQ more than males in the short retest group. There were no Performance

IQ differences with retest condition for females. Males increased more on Full Scale IQ than females in the long term retest group, although both genders showed the retest effect. Gender differences did exist initially on Verbal IQ in the short retest condition and were in the same direction in the long term retest group. It is not clear exactly what these gender differences mean or can be attributed. It is possible that due to the small sample size the groups of males and females may have been atypical. It is also possible there was some gender dynamic occurring between male or female subjects and the two female examiners. However, this has not been supported by previous research (Sattler, 1988) and there is little reason for such an effect if it did occur to influence some IQ indices and not others. Regardless, these results indicate the necessity to continue to investigate the gender factor.

The results of this study are consistent with the proposition that growth in intellectual abilities, as measured by the WAIS-R test of intelligence, continues through the 16 and 17th year, at least for adolescents still attending school.

The clinical implications of this documented trend of growth in intellectual abilities is that premature decisions could be made for individuals on the basis of their measured IQ score at or near 16 years of age. Given the extent to which intelligence tests are utilized within the educational, vocational and clinical realms, professionals need to be aware that adolescents 16 and 17 years of age may still be growing in intellectual ability. Any decisions based on absolute rather than relative IQ scores can be significantly influenced by time of testing (early, late) within the 16.00.00 to 17.11.31 age interval.

As intelligence test norms are based on first time exposure to a test, retest gains need to be understood for 16 and 17 year olds over short and long retest intervals. Given the pronounced influence of practice on measured IQ score, research on the effects of retesting is important to understand predictable gains in measured IQ score due to the effects of practice versus "real" growth. A set of retest norms for different populations and across different retest intervals would provide a standard or base-rate from which gains in IQ scores on retesting could be meaningfully considered (Kaufman, 1990). Test norms which are based on the assumption that an adult level of intelligence is reached by 16 years of age (e.g. Standford Binet) ignores subsequent age differences.

However, given the liklihood of growth in measured IQ beyond 16 years of age, separate standardized norms may be necessary to take into account continued intellectual growth. This is certainly worth considering for 16 and 17 year olds given the results of this study.

Given the significant increase in IQ score in this research on the more educationally related subtests, future research needs to address growth in measured IQ in adolescents no longer attending school. Given the significant difference in initial verbal IQ and some retest IQ indices in this study for males and females, research needs to further investigate the relationship between gender and change/stability in measured IQ.

In conclusion, the results of this study support the contention that growth in intellectual abilities, as measured by the WAIS-R test of intelligence, continues beyond 16 years of age in adolescents still attending school. Subtests more likely to contribute to an increase in IQ score over 18 months are those on the Verbal scale such as Vocabulary and Comprehension which are associated with school and learning (Horn & Cattel's crystallized intelligence). Performance subtests may be more affected by practice and less by actual growth.

Although absolute retest gains were different for the long versus short term retest groups, test-retest reliability was high for both groups. There were some unexpected initial and retest effects by gender which need to be further investigated.

REFERENCES

Archer, R.P.; Marwish, M.; Imhof, E.A.; Piotrowski, L. (1991) Psychological test usage with adolescent clients:1990 survey findings. Professional Psychology:Research and Practice, 22, 247-253.

Bauman, E. (1991) Determinants of WISC-R subtest stability in children with learning difficulties. Journal of Clinical Psychology, 47(3), 431-435.

Bayley, N. (1949) Consistency and variability in the growth of intelligence from birth to eighteen years. The Journal of Genetic Psychology, 75, 165- 196.

Bayley, N. (1957) Data on the growth of intelligence between 16 and 21 years as measured by the wechsler-bellevue scale. The Journal of Genetic Psychology, 90, 3-15.

Bloom, B.S. (1964) Stability and change in human characteristics. New York: John Wiley and Sons, Inc.

Brannigan, G.G. (1975) Scoring difficulties on the wechsler intelligence scales. Psychology in the Schools, 12(3), 313-314.

Brown, H.S. and May, A.E (1979) A test-retest reliability study of the wechsler adult intelligence scale. Journal of Consulting and Clinical Psychology, 47, 601-602.

Catron, D.W. (1978) Immediate test-retest change in WAIS scores amongst college males. Psychological Reports, 43, 279-290.

Catron, D.W. and Thompson (1979) Test-retest gains in WAIS scores after four retest intervals. Journal of Clinical Psychology, 35, 352-357.

Flynn, J.R. (1984) The mean IQ of americans: Massive gains 1932-1978. Psychological Bulletin, 95, 29-51.

Flynn, J.R. (1987) Massive IQ gains in 14 nations: What IQ tests really measure. Psychological Bulletin, 101, 171-191.

Freeman, F.N. and Flory, C.D. (1937) Growth in intellectual ability as measured by repeated tests. _Monographs of the Society for Research in Child Development, 2_ (2, Serial no. 9).

Harrison, P.L.; Kaufmen, A.S.; Hickman, J.A.; Kaufman, N.L. (1988). A survey of tests used for adult assessment. _Journal of Psycho-educational Assessment_, _6_, 188-198.

Horn, J.C. and Cattell, R.B (1966) Refinement and test of the theory of fluid and crystallized intelligence. _Journal of Educational Psychology_, _57_, 253-270.

Jones, H.E. and Bayley (1941) The berkeley growth study. _Child Development, 12_ 167-173.

Kangas, J. and Bradway, K. (1971) Intelligence at middle age: A thirty-eight year follow-up. _Developmental Psychology, 5_, 333-337.

Kaufman, A.S. (1990) _Assessing adolescent and adult intelligence._ Boston: Allyn & Bacon Inc.

Knezevich, S. (1946) The constancy of the IQ of the secondary school pupil. _Journal of Educational Research_, _39_, 506-516.

Matarazzo, J.D.; Carmody, J.P. & Jacops, L.D. (1980) Test-retest reliability and stability of the WAIS: A literature review with implications for clinical practice. Journal of Clinical Neuropsychology, 2, 89-105.

Matarazzo, R.G., Wiens, A.N. Matarazzo, J.D. & Manaugh, J.S. (1973) Test-retest reliability of the WAIS in a normal population. Journal of Clinical Psychology, 29, 194-197.

Matarazzo, R.G. and Herman, A (1980) Test-retest reliability and stability of the WAIS: A literature review with implications for clinical practice. Journal of Clinical Neuropsychology, 2, 89-105.

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

McCall, R.B. Appelbaum, M.L. & Hogarty, P.S. (1973) Developmental changes in mental performance. Monographs of the Society of Research in Child Development, 38, (6, Serial No. 83).

Minton, H.C. and Schneider, F.W. (1980) Differential psychology. Belmont California:Wadsworth Inc.

Moriarty A.E. (1966) Constancy and I.Q. change: A clinical view of relationships between tested intelligence and personality. Springfield, Ill: Charles C. Thomas.

Roediger, H.L, Capaldi, E.D., Paris, S.G. & Polivy, J. (1991) Psychology (3rd edition). New York:Harper Collins.

Ryan, J.J., Prifitera, A. and Powers, L (1983). Scoring reliability on the WAIS-R. Journal of Consulting and Clinical Psychology, 51 (1), 149-150.

Sattler, J.M. (1988) Assessment of children (3rd edition). San Diego:Sattler Publishing.

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

Shatz, M.W. (1981) Comment:WAIS practice effects in clinical neuropsychology. Journal of Clinical

Neuropsychology, 3, 171- 179.

Slate, J.R. and Jones, C.H. (1990). Examiner errors on the WAIS-R:A source of concern. _The Journal of Psychology_, _124_,(3), 343-345.

Terman, L.M. (1917) _The stanford revision and extension of the binet scale for measuring intelligence_. Baltimore:Warwick & York.

Terman, L.M. & Merrill, M.A. (1937) _Measuring Intelligence_. Boston:Houghton-Mifflin.

Thorndike, R.L., Hagen, E.P. and Sattler, J.M. (1986) _Manual for the stanford-binet intelligence scale revised_ (4th edition). Chicago:The Riverside Publishing Co.

Tyler, L.E. (1965) _The psychology of human differences_ (3rd edition). New York:Appleton-Century-Crofts.

Wagner, E.E. & Caldwell, M.S.(1979) WAIS test-retest reliability for a clinical out-patient sample. _Perceptual Motor Skills_, _48_, 131-137.

Warren, S.A. and Brown, W.G. (1973) Examiner scoring errors on individual intelligence tests. Psychology in the Schools, 10, 118-121.

Wechsler, D.C. (1974) Manual for the wechsler intelligence scale for children (Revised). New York:Psychological Corporation.

Wechsler, D.C. (1981) Manual for the wechsler adult intelligence scale (Revised). New York: Psychological Corporation.

Lakehead University

Office of the President

Thunder Bay, Ontario, Canada   Postal Code P7B 5E1

Telephone 343 8200   Area Code 807

November 13, 1990

Dr. A. Thompson and Karen Molly,
Department of Psychology,
Lakehead University,
THUNDER BAY, Ontario.
P7B 5E1

Dear Dr. Thompson:

Based on the recommendation of the Ethics Advisory Committee, I am pleased to grant ethical approval to your research project entitled "To Test the Stability of WAIS-R Scores for 16 and 17 year olds".

Sincerely,

ROBERT G. ROSEHART
President

RGR/uh

THE LAKEHEAD BOARD OF EDUCATION

2135 SILLS STREET
Thunder Bay, Ontario P7E 5T2
Telephone (807) 625-5100
Fax (807) 623-5833

SCHOOLS FOR THE FUTURE

JIM McCUAIG, Director of Education

1990 11 02

Ms. Karen Molly
#211
170 W. Donald Street
THUNDER BAY, Ontario
P7E 5X9

Dear Ms. Molly:

I am pleased to advise you that your research project entitled The Stability of WAIS-R Scores in Sixteen and Seventeen Year Olds has been approved.

A copy of your application and relevant information have been forwarded to the all secondary school principals.

Please contact them directly. Final approval for this research rests with each individual principal. Their decision will be based on factors such as the number of projects in which their school is asked to participate; their opinion of the relevance of the research; and the ~~staff's time considerations.~~

Best wishes for success with your project. This office would appreciate receiving a copy of your report upon completion.

Since~~rely,~~

Curt McMahon
Superintendent of Special Services

CM:ds

pl/3/research.6A

Appendix B

A TEST OF THE STABILITY OF INTELLIGENCE TEST SCORES OVER TIME

DEAR PARENT/GUARDIAN:

I would like your permission to have your son/daughter participate in a research project undertaken by myself, Karen Molly, Masters Student at Lakehead University and Dr. A. P. Thompson, Associate Professor Lakehead University and Registered Psychologist. The research addresses the stability of intelligence test scores. Your child would be asked to participate in two intelligence testing sessions over the next 18 months. Each session lasts approximately 60 minutes. The test is administered individually and in private.

Results of the test will be confidential and not released to school officials. They are to be used solely for research purposes and we are interested in group trends rather than individual scores. However, we are willing to provide your son/daughter with an explanation of their own results after the second session. Your son/daughters participation in this project will reveal valuable information about intelligence testing which has not been thoroughly investigated.

If you approve of your son/daughters participation in this study, please complete the attached consent form and give it to your son/daughter to return to me. If you have any questions or concerns in relation to this research project, please feel free to contact either myself at 473-0786 or Dr. A.P. Thompson at 343-8646. Ethical approval for this research has been received from the Lakehead University Ethics Committee and the Board of Education.

Yours truly,

Karen Molly

Appendix B

PARENT/GUARDIAN CONSENT FORM

I ______________________________ agree to allow my son/daughter

______________________________ to participate in the study on the
(Full Name)

stability of intelligence test scores, conducted by Karen Molly, Masters Student, Lakehead University and Dr. A. P. Thompson, Associate Professor Lakehead University & Registered Psychologist.

I understand that my son or daughters participation will entail being assessed intellectually on two separate occasions, each session running approximately 60 minutes.

Signed _________________________ Date _________________________

Appendix B

A TEST OF THE STABILITY OF INTELLIGENCE TEST SCORES OVER TIME

DEAR PARTICIPANT:

Thank you for agreeing to participate in this research project undertaken by myself, Karen Molly, Masters Student at Lakehead University and Dr. A. P. Thompson, Associate Professor Lakehead University & Registered Psychologist. The research addresses the stability of intelligence test scores. Your participation in this project will reveal valuable information about intelligence testing which has not been thoroughly investigated.

Your participation in this study will involve two (2) intelligence testing sessions over the next 18 months. The tests are administered individually and in private. If you do not anticipate that you will remain a resident of Thunder Bay over this period, please do not volunteer to participate in this study.

Results of the tests will be confidential and not released to school officials. They are to be used solely for research purposes and we are interested in group trends rather than individual scores. However, we are willing to provide you with an explanation of your own results after the second session. Furthermore, if you are interested in the general results of this study, you may request a summary of the findings.

Please sign the attached consent form and return it to me, along with your parents, when we meet. If you have any questions or concerns feel free to contact Dr. A. P. Thompson at 343-8646 or myself, at 473-0786. Ethical approval for this research has been received from the Lakehead University Ethics Committee and the Board of Education.

I will be in contact with you to set up a convenient testing date and time. Once again, thank you for your participation.

Yours truly,

Karen Molly

Appendix B

PARTICIPANT CONSENT FORM

I ______________________________ agree to participate in the study
(Full Name)

on the stability of intelligence test scores, conducted by Karen Molly, Masters student Lakehead University and Dr. A. P. Thompson, Associate Professor Lakehead University & Registered Psychologist.

I understand that I will be assessed intellectually on two separate occasions, each session running approximately 60 minutes.

I understand that all information will be confidential and that I may withdraw my participation in this research project at any time.

SIGNATURE: ____________________________

DATE: ____________________

APPENDIX C: FEEDBACK PROTOCOL FOR SUBJECTS IN THE CONTROL AND EXPERIMENTAL GROUPS

Hello, it's Karen Molly from Lakehead University. I am calling to give you feedback on the test you took. * First I would like to define for you what intelligence is. Basically, it is your ability to solve problems. The intelligence test you took breaks this ability down into two areas:

1) The first is Verbal Intelligence, which is the ability to solve problems using words and numbers.

2) The second area is Performance Intelligence, which is your ability to solve problems without using words and numbers, but rather to solve them visually and often by manipulating things with your hands, for example, the puzzles you did.

**

Now, your particular results were: (we gave feedback only on the first test after the second testing.)

1) Your Verbal Intelligence was in the (score given according to Wechsler Intelligence Classification) range.

2) Your Performance or Nonverbal Intelligence was in the (score given according to Wechsler Intelligence Classification) range.

3) When you put these two results together, your overall problem-solving ability is in the (score given according to Wechsler Intelligence Classification) range.

Do you have any questions? Would you like me to repeat any of this for you?

This is not the absolute or final word on your intellectual level. The scores I have given you are estimates, usually good ones. But the conditions under which you took the test can also influence the results. For example, it you were distracted, nervous or just not trying. I would like to remind you, as well, that Intelligence is only one factor related to success. You need motivation or effort as well. So people with high intelligence can squander their ability and people with lower intelligence can be successful with persistence.

***

If you have any more questions you would like answered, you can call me at 473-0786 and I will make you an appointment with Dr. Thompson at the University.

If anyone in the control or experiment group asked how they did on the second testing, they were told that that is exactly what we were interested in, in this study. But that we are looking at group results not individual, so we were unable to say at that time how people did on average on their second test.

APPENDIX D: FEEDBACK PROTOCOL FOR UNDER 16 YEARS

These subjects received essentially the same feedback. However, at *, the following was added "I will not be retesting you a second time because the study is for people 16 years and older and you were not yet 16 at the time of initial testing". At **, I stated "Before I give you your test results I would like to first caution you. Because the test I gave you was for 16 year olds and up, and you were not yet 16 when you were tested, you took a test designed for older people. Therefore, Dr. Thompson and I feel that your test results likely under-estimate your problem-solving ability. So your test score is possibly lower than if you were to take the test one or two years from now". At *** I added, "Remember, you took a test for older people so your results are probably lower than they would be if you were to take the test one to two years from now".