View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Lakehead University Knowledge Commons

Lakehead University

Knowledge Commons,http://knowledgecommons.lakeheadu.ca

Electronic Theses and Dissertations | Retrospective theses

1994

Comparability of the WAIS-R and WISC-III with 16 year old normals

Sota, Daniela D.

http://knowledgecommons.lakeheadu.ca/handle/2453/888

Downloaded from Lakehead University, KnowledgeCommons

Comparability of the WAIS-R and WISC-III with 16 Year Old Normals

Daniela D. Sota ©
Lakehead University

Thesis Supervisor: Dr. A. P. Thompson

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

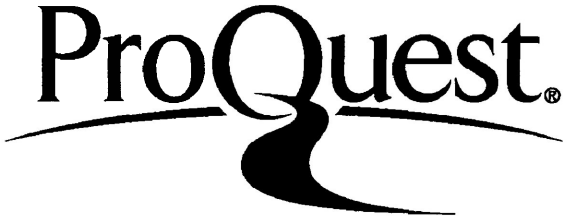
October, 1994

ProQuest Number: 10611417

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.



ProQuest 10611417

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.
This work is protected against unauthorized copying under Title 17, United States Code
Microform Edition © ProQuest LLC.

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

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

THE AUTHOR HAS GRANTED AN IRREVOCABLE NON-EXCLUSIVE LICENCE ALLOWING THE NATIONAL LIBRARY OF CANADA TO REPRODUCE, LOAN, DISTRIBUTE OR SELL COPIES OF HIS/HER THESIS BY ANY MEANS AND IN ANY FORM OR FORMAT, MAKING THIS THESIS AVAILABLE TO INTERESTED PERSONS.

L'AUTEUR A ACCORDE UNE LICENCE IRREVOCABLE ET NON EXCLUSIVE PERMETTANT A LA BIBLIOTHEQUE NATIONALE DU CANADA DE REPRODUIRE, PRETER, DISTRIBUER OU VENDRE DES COPIES DE SA THESE DE QUELQUE MANIERE ET SOUS QUELQUE FORME QUE CE SOIT POUR METTRE DES EXEMPLAIRES DE CETTE THESE A LA DISPOSITION DES PERSONNE INTERESSEES.

THE AUTHOR RETAINS OWNERSHIP OF THE COPYRIGHT IN HIS/HER THESIS. NEITHER THE THESIS NOR SUBSTANTIAL EXTRACTS FROM IT MAY BE PRINTED OR OTHERWISE REPRODUCED WITHOUT HIS/HER PERMISSION.

L'AUTEUR CONSERVE LA PROPRIETE DU DROIT D'AUTEUR QUI PROTEGE SA THESE. NI LA THESE NI DES EXTRAITS SUBSTANTIELS DE CELLE-CI NE DOIVENT ETRE IMPRIMES OU AUTREMENT REPRODUITS SANS SON AUTORISATION.

ISBN 0-315-97071-5

Canada

Table of Contents

Acknowledgements

Very special thanks to my Supervisor, Dr. Anthony Thompson, for his guidance and dedication over the course of this research. Immeasurable gratitude to Patty Ross for her help with testing, as well as her sturdy shoulders and ever-attending ears. The author would also like to acknowledge the valuable comments of Dr. E. Bauman and Dr. J. Jamieson. Finally, I would like to recognize my external examiner, Dr. L. Atkinson.

List of Tables

List of Appendices

Abstract

The comparability, equivalence, and parallelism of the Wechsler Adult Intelligence Scale - Revised (WAIS-R) and the Wechsler Intelligence Scale for Children - Third Edition (WISC-III) was compared using 16 year old youth. A total of 46 subjects, including 16 males and 30 females, were administered the WAIS-R and the WISC-III in a counterbalanced order. The WAIS-R was administered first to 23 subjects, while 23 subjects were administered the WISC-III first, with an average retest interval of 40 days. Repeated measures multivariate analysis of variance (MANOVA) and univariate repeated measures analyses (ANOVA) were used to examine significant test, administration order, and practice effects. The means of the corresponding summary IQs were significantly higher for the WAIS-R than the WISC-III. A clear pattern of practice effects was displayed when the WAIS-R followed the WISC-III. With the exception of two subtests, the means on the WISC-III were significantly higher than the corresponding means on the WAIS-R. The pattern of practice effects for the subtests was ambiguous. Finally, t-tests examined the differences in correlated variance, and Votaw's (1948) test of compound symmetry was employed to compare covariances. The results of these tests showed minor differences. Clinical and theoretical implications of these results are discussed.

Introduction

The most frequently used instruments for measuring intelligence with adolescents are the Wechsler scales (Archer, Mariush, Imhof & Piotrowski, 1991). The scales offer good validity, high reliability, strong standardization, and a wide range of application. However, these scales present unique issues in the assessment of 16 year old subjects. The Wechsler Adult Intelligence Scale - Revised, or WAIS-R (Wechsler, 1981), is intended for use with individuals 16 years and older, while the Wechsler Intelligence Scale for Children - Third Edition, or WISC-III (Wechsler, 1991), has an upper limit of application for individuals aged 16 years, 11 months. This overlap of age ranges requires the clinician to make a choice between the scales when assessing 16 year old subjects. The comparability of scores for 16 year olds on the child and adult versions of the Wechsler scales has been pursued empirically. However, the research literature is quite limited.

Prior Research

Craft and Kronenberger (1979) employed a test-retest design in a counterbalanced order on a sample of 30 educable mentally handicapped students to compare the Wechsler Intelligence Scale for Children - Revised, or WISC-R (Wechsler, 1974), and the Wechsler Adult Intelligence Scale, or WAIS (Wechsler, 1951). The mean retest interval was approximately 37 days. The mean age for all students at the first test was 16.39 years. The results

indicated that the WAIS produced significantly higher scores for Full Scale IQ (FSIQ) than the WISC-R. The mean difference was approximately 12 IQ points (Craft & Kronenberger, 1979). Craft and Kronenberger (1979) did not offer any recommendations for clinicians assessing 16 year olds. The authors did recommend further research on subtest comparability and the appropriate transition age from a child to an adult scale.

A sample of 30 educable mentally handicapped students was also employed to compare the WISC-R and WAIS by Nagle and Lazarus (1979). The tests were administered in a counterbalanced order with a mean retest interval of 10 days. The WAIS yielded significantly higher scores on Verbal IQ (VIQ), Performance IQ (PIQ), and FSIQ. This finding was consistent with that of Craft and Kronenberger (1979). All WAIS subtest scaled scores except Picture Completion were also higher than the WISC-R counterparts. Nagle and Lazarus (1979) suggested that classification systems for individuals of lower ability could be improved by specifying ranges of scores for specific tests. For example, if a WAIS FSIQ in the range of 50 to 70 indicated a mild developmental delay, a WISC-R FSIQ in the range of 46 to 66 may constitute the same classification. The authors did not suggest which test more accurately predicts potential.

Three studies have compared the WISC-R with the WAIS-R. Wechsler (1981) administered the WISC-R and the WAIS-R in a counterbalanced order to 80 normal 16 year olds. The retest interval ranged from one to six weeks. Differences between the

scales for VIQ, PIQ, and FSIQ were 0, 2, and 1 IQ points respectively in favour of the WISC-R (Wechsler, 1981). The WISC-R however, did not produce significantly higher scores, which suggested to Wechsler that the scales were equivalent. Equivalence between the WISC-R and WAIS-R, which had not been found between the WISC-R and the WAIS, may have been related to norm group changes with the newer adult scale. The influence of norms in the equivalence of the Wechsler scales is taken up in a subsequent section. In addition, previous comparisons of the WISC-R and WAIS employed developmentally handicapped individuals, whereas Wechsler's comparison of the WISC-R and WAIS-R employed a normal sample. This difference in samples may also play a role in the discrepant results.

The WISC-R and WAIS-R were further examined using 30 learning disabled students of average intellectual ability (Sandoval, Sassenrath & Penaloza, 1988). Students were administered the scales in a counterbalanced order and the mean retest interval was 5 days. Although 5 WISC-R subtest scaled score means were significantly higher than corresponding WAIS-R subtest means, there was no significant difference between the WISC-R and WAIS-R FSIQ scores. Verbal and Performance IQs were not compared. The Sandoval et al. results are consistent with Wechsler's (1981) findings.

A substantive examination of WISC-R and WAIS-R comparability was undertaken by Quereshi, Treis and Riebe (1989). A sample of 144 randomly selected 16 year old high school students were

administered the scales in counterbalanced order with a mean retest interval of approximately 25 days. The mean age for all students at the first test was 16.5 years. The study was conducted in two phases with 72 students tested in each phase. For the first group of 72 students, the WISC-R yielded significantly higher scores than the WAIS-R, with the exception of VIQ and Picture Arrangement. For the second group of 72 students, the WISC-R scores were again significantly higher than the WAIS-R scores, except VIQ and Digit Span subtest scaled score means (Quereshi, Treis & Riebe, 1989). These findings contradict the results of Wechsler (1981) and Sandoval, Sassenrath and Penaloza (1988). Quereshi, Treis and Riebe (1989) suggested that non-equivalence of the scales was a result of differences in the number of subtest items, their content, and scaled score equivalents. This item is addressed in a subsequent section. The authors conclude that it was unclear which scale might more accurately predict the potential performance among 16 year olds. However, Quereshi et al. did caution administrators that the scales should not be assumed to provide the same information in clinical or educational assessment of 16 year old adolescents.

Most recently, Wechsler (1991) employed 189 normal 16 year olds to compare the WISC-III and WAIS-R. The scales were administered in a counterbalanced order with a mean retest interval of 21 days. The WAIS-R FSIQ, VIQ and PIQ were 4, 2 and 6 points higher than the corresponding WISC-III scores (Wechsler, 1991). It was not clarified whether these differences were

significant. However, high correlations suggested to Wechsler that the two scales were measuring similar constructs. Comparative data on corresponding subtests was provided, although Wechsler (1991) cautions against direct comparison of subtest scaled scores because of reference group differences. Wechsler (1991) recommended that clinicians employ the WISC-III to assess lower functioning 16 year olds because the scale provided finer discrimination and extended to a lower IQ range than the WAIS-R.

Table 1 summarises the empirical work which has compared the adult and child versions of the Wechsler intelligence scales with 16 year old individuals. Some consistencies emerge in the findings, however practitioners need to appreciate the following issues in attempting to integrate the research.

Comparability

The terms comparable, equivalent and parallel are not synonymous (Quereshi, Treis & Riebe, 1989). Two measures can yield comparable scores even when they measure different psychological constructs. Thus, a measure of mathematical reasoning and a measure of verbal comprehension produce comparable results when an individual scores at the 75th percentile on both measures. Quereshi, Tries, and Riebe (1989) maintain that the measurement and evaluation literature, although inconsistent, generally uses the term equivalence for alternate test forms or measurements which have the same true scores. Thus, equivalent measures have equal population means or equal

Table 1

Prior Research Comparing the Wechsler Adult and Child Intelligence Scales

Study	Sample[1]	Age	Retest[2]	Tests	Results
Craft & Kronenberger (1979)	30 EMH	16.4	37	WAIS/ WISC-R	WAIS FSIQ > WISC-R FSIQ
Nagle & Lazarus (1979)	30 EMH	16-0 - 16-11	10	WAIS/ WISC-R	WAIS P/V/FSIQ > WISC-R P/V/FSIQ WAIS subtests > WISC-R (except PC)
Wechsler (1981)	80 N	16-0 - 16-11	7 - 42	WAIS-R/ WISC-R	WISC-R FSIQ = WAIS-R
Sandoval, Sassenrath & Penaloza (1988)	30 LD	16-0 - 16-11	5	WAIS-R/ WISC-R	WISC-R FSIQ = WAIS-R 5 WISC-R subtests > WAIS-R
Quereshi, Treis & Riebe (1989)	144 N	16.5	25.4	WAIS-R/ WISC-R	WISC-R > WAIS-R (except VIQ/ DSp/PC)
Wechsler (1991)	189 N	16-0 - 16-11	21	WAIS-R/ WISC-III	WAIS-R > WISC-III

[1]EMH = educable mentally handicapped, N = normals, LD = learning disabled.

[2]measured in days.

means in any probability sample from the population. However, the terms equivalent and parallel are also not interchangeable. Equivalent measures have equal mean scores. Yet, equivalent measures may differ greatly in the range of scores achieved. In addition to equal true scores, parallel measures have equal variances and covariances (Lord & Novick, 1968).

Quereshi, Treis and Riebe (1989) have been the only researchers to investigate the parallelism of the adult and child versions of the Wechsler scales. Cumulative grade point average and a general ability test were employed as criterion measures to compute covariance. It was found that the WAIS-R and WISC-R, except for VIQ, were neither equivalent nor parallel for FSIQ. All other research comparing Wechsler scales has been limited to comparisons of subtest or summary IQ score means (i.e., investigating the equivalence of the scales, as in Craft & Kronenberger, 1979; Nagle & Lazarus, 1979; Wechsler, 1981; Sandoval, Sassenrath & Penaloza, 1988; Wechsler, 1991). It seems important to investigate both the equivalence and the parallelism of the WAIS-R and the WISC-III.

Norms

Over the past 15 years, Wechsler scales that have been compared were normed a number of years apart. Greater comparability might be expected if the adult and child versions had more comparable chronological norms. For example, the WAIS and WISC-R were normed 19 years apart, with the WISC-R being the

more recent test. The older WAIS produced significantly higher scores than the WISC-R in the two empirical studies previously reviewed. The norming of the WISC-R and the WAIS-R was seven years apart, with the WAIS-R being the more recent test. Two of the three studies comparing these scales found no significant differences between the scales (Sandoval, Sassenrath & Penaloza, 1988; Wechsler, 1981). Finally, the norming of the WAIS-R and the WISC-III was ten years apart, with the WISC-III being the more recent test. The WAIS-R seems to produce higher scores than the WISC-III (Wechsler, 1991). Possibly, the more divergent the norms (ie. 10 years or more), the less likely the results for 16 year olds will be equivalent. Also, when differences have been found between the adult and child versions of the Wechsler scales, the scale that had the older norms produced significantly higher IQ scores (Craft & Kronenberger, 1979; Nagle & Lazarus, 1979; Quereshi, Treis & Riebe, 1989; Wechsler, 1991).

In addition to a general pattern of non-equivalence related to the chronology of norming, anomalies with specific versions of the Wechsler scales may detract from equivalence. For example, 16 and 17 year old individuals in the WAIS-R norm sample performed substantially lower than the same age group used for the norming of the WAIS, as demonstrated by decreased norms on the WAIS-R (Kaufman, 1990). This implies that IQs have decreased over time. However, Flynn (1984) has determined that IQs have been steadily increasing. Hence, WAIS-R norms for 16 year olds seem to incorporate some type of unknown bias limiting the norms.

Kaufman (1990) has implicated the procedures that were employed to select the 16 to 19 year old and the 20 to 74 year old individuals for the WAIS-R standardization sample. For adolescents, the occupation of the subject's head of the family was employed for stratification. However, for the adults, their own occupation was used (Wechsler, 1981). If the correlation between the occupation of the family head and the future occupation of the adolescent offspring is weak, stratification using the WAIS-R methods would provide a biased sample. The true basis for the bias is uncertain. Kaufman (1990) has suggested that scaled scores for 16 to 19 year olds obtained using the WAIS-R should be interpreted with caution. Kaufman (1990) has further concluded that in general, for clinical purposes, reason prescribes using the instrument with the more recent norms when assessing 16 year olds.

Sample

Prior research investigating the equivalence of the Wechsler scales has employed a variety of samples including educable mentally handicapped and learning disabled students, as well as normal subjects. From a clinical perspective, level of functioning may influence which Wechsler scale is chosen for assessing 16 year olds. For individuals in the lower range of IQ, the WISC-III could be the scale of choice as a result of its ample "floor" for items and scales. Wechsler (1991) suggests assessment of lower functioning individuals with the WISC-III, as

it was designed for younger children, and may provide more accurate appraisal of ability. On the other hand, for a 16 year old individual of exceptional ability, the WISC-III may not provide an ample "ceiling". The WAIS-R may be preferred in this instance as it would provide a greater range.

For Atkinson (cited in Sattler, 1992), level of functioning is less of an issue when assessing 16 year olds. He has pointed out that the WISC-III, in comparison to the WAIS-R, has better subtest reliabilities, lower subtest floors, better item gradients below the mean, a lower FSIQ floor, and a higher FSIQ ceiling. Therefore, he concluded that with 16 year olds, regardless of the level of intellectual functioning, the WISC-III should be employed for assessment. However, the issue of test choice based upon level of functioning is controversial. For most 16 year old individuals who fall between the upper and lower ranges of IQ, there is little information to guide the clinician about which Wechsler scale to employ. When placement is an issue, individuals functioning at the same level may be placed in different educational settings solely on the basis of the assessment instrument chosen (Rubin, Goldman & Rosenfeld, 1985). Clearly data on the equivalence of the scales would be of great relevance for practitioners and their clients.

Item Content and Scoring

The WAIS-R and WISC-III are composed of ten nomologically similar subtests. However, only Digit Symbol/Coding and Digit

Span can be considered entirely analogous in content. The Comprehension subtest on the WISC-III is designed to measure an adolescent's knowledge of interpersonal relations and social mores (Sattler, 1992). However, there are three questions on the Comprehension subtest of the WAIS-R that require the examinee to understand the proverbial sense of words. Clearly these items deviate from all questions on the WISC-III Comprehension subtest and the majority of questions on the WAIS-R Comprehension subtest (Quereshi, Treis & Riebe, 1989). As well, the WISC-III includes an optional subtest, Symbol Search, not included in the WAIS-R.

In addition to subtest content differences, scoring is different for the two scales. The WISC-III includes answers with both 1 and 2 point maximums for the subtest Similarities, while the WAIS-R has only responses with two point maximums. Although the number of items for the subtest Information on the WAIS-R and the WISC-III differs by only one, the raw score equivalents of a scaled score of 10 for each of the two scales differs by three points. Other differences are also present. Picture Arrangement on the WISC-III includes bonus points for quick performance, while the WAIS-R does not have this feature. An administrative difference between the test occurs for Object Assembly, where on the WISC-III, the object to be assembled is revealed to the examinee for the first two items. This is not included for any of the items in the WAIS-R. Differences also exist when obtaining scaled scores for the subtests. On the WAIS-R, the subtest raw score is procured by comparing the raw score to the

raw scores of all other age groups. The scaled score is then converted to a summary IQ by comparison with other 16 year olds. On the WISC-III, the scaled score is computed by comparing the raw score with the scores of other 16 year olds, then the summary IQ is obtained by comparison of the scaled score to all other age groups. With such differences between the scales, equivalence/parallelism might not be expected between the subtests and summary IQs.

Practice Effects

In order to determine the equivalence of the Wechsler scales, both tests must be administered. Many have investigated the effect of readministering a Wechsler scale after a short interval (Catron, 1978; Catron & Thompson, 1979; Matarazzo, Wiens, Matarazzo & Manaugh, 1973; Wechsler, 1991). Upon retest, subjects may be less anxious and less distracted as a result of familiarity with the test (Matarazzo, Wiens, Matarazzo & Manaugh, 1973). In addition, it has been proposed that the formation of a strategy for subtests such as Block Design facilitates solving designs more capably upon retest (Catron, 1978). Collectively, these are referred to as practice effects. Thompson and Molly (1993) found that for 16 year olds, retest effects vary with the length of the retest interval. The authors found that gains in IQ could be attributed to practice effects over a three month retest interval. However, gains in IQ over an 18 month retest were more than practice effects and attributed to

continued intellectual growth in this age range.

Practice effects can be controlled for by counterbalancing administrations and by maintaining similar retest intervals for all subjects. Short retest intervals less than three months can also minimize retest effects from intellectual growth among 16 year olds. All studies in Table 1 comparing the adult and child versions of the Wechsler scale have counterbalanced administration order to control practice effects and have used short retest intervals.

Examiner Error and Situational Variables

Studies have investigated the prevalence and consequences of examiner scoring errors. Errors are in fact common among both certified psychologists as well as students (Sattler, 1988; Slate & Jones, 1990; Slate, Jones, Coulter, & Covert, 1992). A small percentage of computational errors can be very large (Thompson & Hodgins, in press). Examiners were found to make as many as 11 errors per protocol (Slate & Jones, 1990), and as many as 88% of FSIQs were affected (Slate, Jones & Murray, 1991). Particularly prone to error were the Verbal subtests Vocabulary, Comprehension, and Similarities (Slate & Chick, 1989). Examiners often assigned too few or too many points to answers, failed to query, or questioned inappropriately (Slate, Jones, Coulter, & Covert, 1992). Slate and Hunnicutt (1988) attribute such errors to poor instructional preparation and ambiguity or paucity of manual scoring criteria. Even Wechsler subtests with objective

scoring criteria, such as Coding and Digit Span, can be affected by carelessness, failure to record either responses or times, or the acquisition of incorrect basals or ceilings (Slate & Hunnicutt, 1988). Clearly, the reliability and validity of scores on the Wechsler tests can be decreased by examiner errors.

It has also been demonstrated that situational variables can influence performance and scoring. Sattler and Gwynne (1982) reviewed 29 published studies investigating examiner race effects on individual intelligence tests and other cognitive measures. In the majority of studies, there was no significant relationship between race of examiner and examinee performance. However, subjects may display superior performance with female, as opposed to male examiners. The research is not conclusive on this issue (Sattler, 1988). Fuchs and Fuchs (1986) found elevated test scores of 7.6 points or more on intelligence test when the examinee was familiar with the examiner. Familiarity consisted of previous acquaintance or prior contact with the examiner. Finally, pretest information, including case history, prior test scores, grades, ethnicity, or sex, can influence the scoring of responses, especially when responses are ambiguous (Sattler & Winget, 1970). Thus, elements of subjectivity have been confirmed by research and precautions must be taken to avoid them. Sattler (1988) recommended that examiners must remain fair and consistent, and reduce sources of subjectivity or error.

Current Investigation

The present research used a counterbalanced test-retest paradigm to investigate the equivalence of the WAIS-R and the WISC-III for 16 year old adolescents. A measure of achievement was also obtained in order to examine the parallelism of the scales. Test-retest intervals were short and consistent to minimize the effects of mental growth among individuals. Standardized administration and accurate scoring were emphasized.

Method

Subjects: Forty-eight subjects were recruited from five local high schools in the public and separate school systems. Complete data (test and retest) from 46 subjects made up the final sample for the study. The unused subject data resulted from a misplaced protocol in one instance, and questionable subject motivation in the other instance. The mean age of subjects at first administration was 16 years, 4.9 months with a range of 16 years, 0 months, 2 days to 16 years, 10 months, 16 days. Sixteen of the subjects were male, thirty were female. All subjects were Caucasian except one subject of Oriental ethnicity.

Measure: The Wechsler Adult Intelligence Scale - Revised (Wechsler, 1981) and the Wechsler Intelligence Scale for Children -Third Edition (Wechsler, 1991) were administered to all

subjects. All subtests were administered except the supplementary WISC-III subtests (Mazes, Symbol Search, Digit Span). As well, each subject was asked to recall the percent average (PA) that they had achieved in the previous academic year.

Procedure: Ethical approval for the research was obtained from the Lakehead University Ethics Advisory Committee (Appendix A). Approval was also granted by the Lakehead School Board and the Lakehead Separate School Board (Appendix A). All subjects reported that they had not been administered any intelligence scale in at least four years. A Consent to Participate form was signed both by the participant and the parent or guardian (Appendix B).

Two female administrators at the M.A. level conducted all testing according to standardized procedures. Both examiners had completed a graduate course in psychometric assessment which included instruction with either the WAIS-R or WISC-III, and a competency test. Test administration was further reviewed and practised with three practitioners experienced with the Wechsler scales. Administrators were alerted to common administration and scoring errors (Slate & Jones, 1990).

Subjects were tested as they were recruited. Assignment of subjects to administrators was based upon administrator availability. Order of test administration was by random lot. The WAIS-R was administered first to 15 females and 8 males,

while the same number of subjects by sex were administered the WISC-III first. The subjects by order of administration did not differ significantly in mean age ($t(44)=1.23$, $p = .64$) or mean retest interval ($t(44)=1.07$, $p = .88$). The mean retest interval was 40.0 days, with a range of 29 to 48 days. All subjects were tested and retested by the same administrator. The distribution of subjects by examiner was 27 versus 19 and the breakdown of subject sex by examiner was 9 male and 18 female subjects versus 7 male and 12 female subjects. During either the first or second testing session, subjects were asked to provide a specific estimate of their year-end percent average from the previous academic year. After the second testing session, each subject was provided feedback from their first test according to a standard protocol (Appendix C).

Examiners scored their own protocols. Protocols of both examiners were double checked for computational and clerical errors by an experienced administrator. To ensure consistency of scoring, this individual also rescored the Verbal subtests most open to examiner judgement (i.e., Vocabulary, Comprehension, Similarities). Scoring revisions were made as necessary.

Treatment of Data: The primary data were obtained from 92 protocols and consisted of six IQ estimates and 21 subtest scaled scores for each of the 46 subjects. Self-reported percent average (PA) served as an external validity criterion.

To examine the equivalence of means, the 10 subtest scaled

scores and 3 IQ estimates were subjected to a repeated measures multivariate analysis of variance (MANOVA), as well as a univariate analysis of variance (ANOVA). In these analyses, counterbalanced administrative order (WAIS-R/WISC-III versus WISC-III/WAIS-R) was the between-subjects factor, and the test (WAIS-R versus WISC-III) was the within-subjects factor. The interaction term was used to examine practice effects.

Intertest correlations (Pearson *r*) for corresponding subtest scaled scores and corresponding summary IQs were calculated. To examine equivalence of variance, t-tests for correlated variance were performed for each of the corresponding Wechsler subtests and summary IQs.

Correlation coefficients and covariances between Wechsler scores and the PA were computed. To determine the equivalence of relationships between the Wechsler scales and a criterion, Votaw's (1948) test of equivalence of variance and covariance, further delineated by Gulliksen (1950), was performed. Votaw's test statistic L_{vc} is defined as

$$L_{vc} = \frac{s_y^2 \, s_1^2 \, s_2^2 \, [\, 1 + 2r_{y1} \, r_{y2} \, r_{12} - r_{y1}^2 - r_{y2}^2 - r_{12}^2 \,]}{[\, s_y^2 \, (u_x + w_x) - 2c_{yx}^2 \,][\, u_x - w_x \,]} ,$$

where, s^2 = variance

r = correlation

c = covariance

y = criterion (PA)

1 = order 1 (WAIS-R/WISC-III)

2 = order 2 (WISC-III/WAIS-R)

$u_x = (s_1^2 + s_2^2) / 2$

$w_x = c_{12}$

$c_{yx} = (c_{y1} + c_{y2}) / 2$

Results

Comparison of Means

Means and standard deviations for subtests and summary IQs for each administrative condition are presented in Table 2. Means and standard deviations collapsed across administrative order are also provided. The present sample was generally in the average to high average range of intelligence. WAIS-R FSIQ ranged from 81 to 136 and WISC-III FSIQs ranged from 84 to 137. Intertest correlations were calculated for each administrative order and then combined by correcting for variability with Fisher's z transformations. For Verbal subtests, intertest correlations ranged from .45 (Comprehension) to .79 (Vocabulary). Performance subtest intercorrelations ranged from .16 (Picture Arrangement) to .73 (Coding). Finally, summary IQs correlated .88 (FSIQ), .85 (VIQ), and .74 (PIQ).

It was concluded that the data complied sufficiently with MANOVA assumptions (Tabachnick & Fidell, 1989) to justify the procedure. Specifically, there were more subjects than dependent variables (DVs) per cell of the design. It seemed reasonable to

Table 2

Mean and Standard Deviation (in Parentheses) for Wechsler Subtests and Summary IQs[1]

Order[2]	1		2		Combined	
Session	1	2	1	2	1	2
Test	WAIS-R	WISC-III	WISC-III	WAIS-R	WAIS-R	WISC-III
Info	7.7(2.4)	10.2(3.4)	11.0(2.3)	8.5(2.3)	8.1(2.4)	10.6(2.3)
DSp	8.9(1.9)	---	---	9.9(1.9)	9.4(1.9)	---
Vocab	9.4(1.7)	10.7(1.7)	10.8(1.7)	9.4(2.1)	9.4(1.9)	10.7(1.7)
Arith	9.1(2.3)	10.7(3.1)	9.6(2.9)	9.9(2.6)	9.5(2.4)	10.2(3.0)
Comp	9.7(2.4)	10.9(1.9)	11.6(2.0)	10.2(1.8)	9.9(2.1)	11.2(2.0)
Sim	9.9(2.3)	11.5(2.4)	11.6(2.7)	10.8(1.8)	10.3(2.1)	11.5(2.5)
PC	9.7(1.9)	12.2(2.6)	12.0(2.6)	11.2(1.7)	10.4(2.0)	12.1(2.5)
PA	10.7(2.9)	12.4(3.4)	11.3(3.1)	10.5(2.3)	10.6(2.6)	11.8(3.3)
BD	11.0(2.9)	11.4(3.1)	10.9(2.7)	12.7(2.2)	11.9(2.7)	11.2(2.9)
OA	11.3(3.2)	12.0(3.1)	11.9(3.5)	12.1(2.4)	11.7(2.8)	11.9(3.2)
Cod	11.4(3.0)	13.3(3.8)	11.6(2.3)	12.6(1.7)	12.0(2.5)	12.4(3.2)
VIQ	104.7	104.8	105.5	109.3	107.0	105.2
	(10.5)	(10.9)	(10.0)	(9.3)	(10.1)	(10.3)
PIQ	111.0	115.0	110.5	120.4	115.7	112.8
	(17.1)	(17.0)	(11.1)	(8.9)	(14.3)	(14.4)
FSIQ	108.1	110.6	108.4	114.9	111.5	109.5
	(13.4)	(13.6)	(9.2)	(8.5)	(11.6)	(11.6)

[1]WAIS-R scaled scores are based upon 20-34 year old reference group; WISC-III scaled scores are relative to same age reference group.

[2]order 1 = WAIS-R/WISC-III; order 2 = WISC-III/WAIS-R.

assume that DVs were normally distributed in the population and sample size greater than 20 subjects ensured robustness. The cell data were examined for outliers, defined as a standard score more extreme than ± 3, and none were found. Variances of the DVs in "between-subjects" cells of the design were comparable and the ratio of smallest to largest was well within the criterion suggested by Tabachnick and Fidell (1989). Homogeneity of variance-covariance matrices was assumed based upon equal cell sample sizes. Linear relationships between all pairs of DVs was assumed but not checked. Hence, some statistical power may have been lost if the data departed from linearity. Finally, the DVs were in some cases highly correlated, but no coefficients exceeded .90. Although summary IQs were combinations of other DVs, singularity and multicollinearity were not flagged as present by the SPSS programme. In addition, Pillai's F statistic was used as the criterion for the multivariate tests as this statistic is robust to violations of MANOVA assumptions while having good power.

A repeated measure MANOVA was conducted on the 13 repeated measures (i.e., 10 subtest scaled scores and 3 summary IQs). The multivariate main effect for administrative order was $F(13,32) = 3.47$, $p = .002$; the multivariate main effect for tests was $F(13,32) = 7.36$, $p < .001$; and there was a significant multivariate interaction, $F(13,32) = 3.40$, $p = .002$.

The significant multivariate effects were explored further through univariate repeated measures ANOVAs. The univariate test

effect was significantly different for Information (p <.001), Vocabulary (p <.001), Comprehension (p <.001), Similarities (p <.01), Picture Completion (p <.001), Picture Arrangement (p <.05), and Block Design (p <.05). In addition, Arithmetic (p <.06) showed differences in means that were close to significant. All of the subtest scaled score means were larger on the WISC-III than the WAIS-R except Block Design which was larger on the WAIS-R. The univariate test effects for VIQ ($F(1,44) = 5.01$, p =.030), PIQ ($F(1,44) = 3.76$, p =.059) and FSIQ ($F(1,44) = 5.20$, p =.028) were significant with WAIS-R summary IQs greater than WISC-III summary IQs. The following effect sizes were calculated: VIQ (eta-squared = .008), PIQ (eta-squared = .01), FSIQ (eta-squared = .007). These effect sizes were very small.

Univariate interaction effects were significant for Arithmetic (p <.01), Picture Completion (p <.05), Block Design (p <.01), and Coding (p <.001). All three summary IQs revealed an order by test interaction: FSIQ, $F(1,44) = 25.56$, p <.001; VIQ, $F(1,44) = 5.25$, p =.027; and PIQ, $F(1,44) = 21.62$, p <.001. The significant interactions were explored by examining simple effects with paired t-tests. Using the .05 significance level, these tests revealed no clear pattern of practice effects for the subtests. Thus, Arithmetic and Picture Completion showed a significant practice effect only when the WISC-III followed the WAIS-R. For Block Design, a practice effect occurred only when the WAIS-R followed the WISC-III. The Coding subtest showed a significant increase in mean scaled score on the second

administration regardless of whether it was the WAIS-R or the WISC-III. Paired t-tests with summary IQs revealed that a significant practice effect ($p < .01$) existed only when the WAIS-R followed the WISC-III and not vice versa.

Comparison of Variances and Covariances

Tests for correlated variance were conducted for each group of subjects by administrative order and for the combined sample. For the group taking the WAIS-R first, no differences in variance were found for any of the subtests or summary IQs. For the group taking the WISC-III first, a significant difference in variance was found for Picture Completion, $t(22) = 2.24$, $p < .05$, and Object Assembly, $t(22) = 2.09$, $p < .05$. For the combined sample, the difference between correlated variances for the subtests Arithmetic, $t(44) = 2.05$, $p < .05$, and Coding, $t(44) = 2.35$, $p < .05$, were significant.

Finally, variance, covariance, and correlation coefficients were used in a test of compound symmetry to examine the equality of validity coefficients. The subjects' estimated PA, which served as the criterion, ranged from 50 to 93 percent with a mean of 76.8 percent. Table 3 presents correlations and covariances between the subtests, IQ scores and the PA. Application of Votaw's (1948) test produced 13 chi-squares, one for each of the 10 subtests and 3 summary IQs. Only the differences in covariance for Coding were found to be significantly different (χ^2 (2, N=46) = 3.48, $p < .05$).

Table 3

Correlation and Covariance Between Subtest Scaled Scores, Summary IQs and Percent Average (PA)

	Correlation with PA		Covariance with PA	
	WAIS-R	WISC-III	WAIS-R	WISC-III
Information	.23	.46	6.26	12.54
Vocabulary	.38	.52	8.37	10.17
Arithmetic	.45	.33	12.81	11.56
Comprehension	.33	.34	8.07	7.76
Similarities	.33	.23	8.13	6.86
Picture Completion	-.004	.10	-.10	2.83
Picture Arrangement	.03	.31	1.01	11.69
Block Design	.36	.37	11.26	12.41
Object Assembly	.21	.28	7.07	10.48
Coding	.52	.35	15.09	13.17
FSIQ	.49	.50	65.26	67.67
VIQ	.50	.48	57.87	57.36
PIQ	.35	.40	58.31	66.91

Discussion

The results of this study showed that the WAIS-R and the WISC-III differed on a number of psychometric properties when compared in a counterbalanced order with 16 year old students. However, some of these differences were predictable or of limited degree. The results revealed a closer correspondence between the WAIS-R and WISC-III than was found in a previous study that compared the WAIS-R and WISC-R (Quereshi, Treis & Riebe, 1989). The results also showed some areas of closer correspondence between the WAIS-R and WISC-III than reported in the WISC-III manual. Specific results are discussed in the following.

First, intertest correlations for the subtests and summary IQs were comparable to those reported in a WAIS-R/WISC-III comparison study by Wechsler (1991). Wechsler considered the magnitude of correlations to be sufficiently high to conclude that the two tests measured similar constructs. The same could be said of the intertest correlations from this study with the exception of Picture Arrangement. Wechsler found Picture Arrangement to have the lowest intertest correlation (r = .45), but it was considerably higher than for the current study (r = .16). Both subtests measure similar non-verbal reasoning skills. However, the WISC-III introduces speed as a factor with up to three bonus points on most items for rapid execution. There are no bonus points for speed on WAIS-R Picture Arrangement, and this difference is the most likely explanation for poor correlation

between the two measures.

Second, half of the WISC-III subtest scaled scores were significantly higher than corresponding WAIS-R subtest means. Quereshi et al. (1989) found, for each of two groups, that all but one subtest had significantly higher WISC-R means compared with WAIS-R means. Differences in subtest means may be related to differences between the two tests in obtaining scaled scores. For the WAIS-R, scaled scores are relative to a 20-34 year old reference group. Scaled scores on the WISC-III are relative to a same age reference group. For this reason, Wechsler (1991) cautions that direct comparisons of subtest scaled scores should not be made. However, if reference group differences is a factor in the discrepancy, it is not evident across all subtests. To further explore subtest comparability, it would be valuable to convert WAIS-R raw scores to age-related scaled scores before comparisons were made.

Third, mean summary IQs were significantly higher on the WAIS-R than on the WISC-III. This finding is in accordance with prior research in which the most recently normed scale provided lower IQ scores. This may be a result of IQs steadily increasing over time (Flynn, 1984). Hence, more recently normed Wechsler scales compare an examinee to a group with higher IQ norms. Thus, applying the results of this study to the definition of equivalence (i.e. equal means), the child and adult versions of the Wechsler intelligence scales cannot be considered equivalent. This finding is consistent with most research over the past 15

years comparing Wechsler scales for 16 year olds. Although significant differences between the summary IQ means were found, the magnitude of the differences were smaller than those found by Wechsler (1991). Wechsler (1991) found VIQ, PIQ, and FSIQ differences of 2, 6, and 4 points in favour of the WAIS-R. In this study, corresponding differences were 2, 3, and 2 IQ points. Indeed, these results reveal a more encouraging depiction of the comparability of means. Moreover, measurement error at a 68 percent confidence level for 16 year olds on the WISC-III is $\pm$3.35, $\pm$4.24, and $\pm$3.00 for VIQ, PIQ, and FSIQ respectively (Wechsler, 1991). Hence, mean differences between the WAIS-R and the WISC-III may not be clinically significant. As it is typical to report an IQ range rather than a specific IQ number, results from the WAIS-R and the WISC-III would be expected, in many cases, to show considerable overlap. In addition, it was concluded from the effect size statistic that test (WAIS-R versus WISC-III) accounted for a very small percent of the total variance of summary IQs. This supports the contention that the differences between the WAIS-R and the WISC-III at the level of summary IQs may not be clinically significant.

Fourth, the results revealed no clear pattern of subtest practice effects when both the WAIS-R and WISC-III are administered. However, clinicians need to remain aware of this possibility, particularly it would seem with Performance subtests. In this study, Block Design, Picture Completion, and Coding were most likely to show practice effects related to

administration order. Kaufman (1990) and Sattler (1992) advise that Performance subtests are most prone to practice effects. Indeed, three of the four subtests that demonstrated practice effects were Performance subtests. For summary IQ, there was a clear pattern associated with practice when the WAIS-R followed the WISC-III, but not vice versa. Summary IQs were, on average, considerably larger (4 to 10 IQ points) when the WAIS-R followed the WISC-III. This difference is a combination of practice and differences between the two scales. When the WISC-III followed the WAIS-R, practice effects generally seemed to counteract the typically higher WAIS-R IQs.

Finally, comparisons of variance revealed diverse differences according to group. For the combined group, only 2 of the 10 subtests were found to have significantly different variances. The group administered the WAIS-R first demonstrated no differences in variance, and the group administered the WISC-III first demonstrated differences in variance on 2 of the 10 subtests. These subtests, however, were dissimilar to those exhibiting significant differences in the combined sample. Votaw's test of compound symmetry revealed that only one subtest displayed a significant difference in covariance. Consequently, according to definition of parallelism, the WAIS-R and the WISC-III are not parallel. However, Quereshi, Treis and Riebe (1989) caution that the authors and publishers of the Wechsler scales never purport parallelism between the scales. Thus, it may be unjustifiable to apply the criteria of parallelism. Nonetheless,

these results are again more promising than those of Quereshi et al., who found differences in variance for as many as seven of the subtests, and differences in covariance for the same number of subtests and one summary IQ score.

The results of this study, although consistent with prior research, reveal smaller differences in subtest and IQ means, variance, and covariance for the child and adult versions of the Wechsler scales. This may be in part a result of the employment of meticulous scoring procedures. Each protocol was scored once by the administrator, then clerical items and calculations were rechecked by an experienced practitioner. Finally, the Verbal subtests most subject to judgement errors were rescored by the same practitioner to ensure consistency. Nevertheless, a much smaller sample was employed in this study, as compared with Wechsler (1991) and Quereshi, Treis, and Riebe (1989). Although the sample was statistically large enough, it may be that the larger samples were more representative. Another limitation of this study was the potential for inaccuracy in the self-reported percent average. However, this method was the most expedient route to acquire the percent average as a result of the obstacles involved in obtaining confidential student records. Subjects appeared to have little difficulty recalling their grades from the preceding year, and were queried if they gave a range of five percentage points or greater. Thus, the self-report seemed reasonable.

Wechsler (1991) has suggested that practitioners employ the

WISC-III with 16 year olds of lower intellectual functioning because of its capability of finer discrimination and lower subtest scaled score basals. Furthermore, Atkinson (cited in Sattler, 1988) advises that the WISC-III be used with all 16 year olds. Theoretically, the differences in scores between the WAIS-R and the WISC-III are statistically significant. However, they may not be clinically significant. Thus, practical implications of this research are that the summary IQ means on the WAIS-R and the WISC-III are not radically incongruent for 16 year olds of normal intellectual functioning. Nevertheless, this does not negate Atkinson's recommendation for utilization of the WISC-III with all 16 year olds.

Clearly, in the field of psychology, there is great discrepancy in the use of the terms comparable, equivalent, and parallel. Branch (1990) emphasizes that two tests do not necessarily measure the same construct, nor are scores interchangeable as a result of a high correlation. Means, variances, and item content must be scrutinized (Branch, 1990). Indeed, Nagle and Lazarus (1979) and Craft and Kronenberger (1979) report that the WAIS and WISC-R appear to be highly related as a result of significant correlations. Fine (1992) subsequently elaborated on this point to make educational and psychological practitioners cognizant of the distinction between tests that measure the same construct, and parallel forms of tests. For tests measuring the same construct, convergent validity, or correlations with tests measuring theoretically

related constructs, should be high, while there should be low correlation with tests that measure unrelated constructs, or divergent validity (Fine, 1992). Furthermore, one must discriminate between parallel forms of tests and statistically parallel tests. Parallel or alternate forms designate equivalent measures, or measures with equal means and variances. The term statistically parallel encompasses tests which measure similar constructs, have similar item content, and have equal means, variances, and covariances. Because distinctions can be made between the terms, great care should be utilized with employing them. As well, practitioners must be made aware of the discrepancies. Future research could include further investigation into statistical parallelism to make practitioners aware of the differences between the terms, as well as the scales themselves. Prospective consideration should also involve research into the appropriate age of transition from a children's version to an adult version of a scale, perhaps with no overlap in ages.

References

Archer, R. P., Maruish, M., Imhof, E. A., & Piotrowski. (1991). Psychological test usage with adolescent clients: 1990 survey findings. Professional Psychology: Research and Practice, 22, 247 - 252.

Branch, W. (1990). On interpreting correlation coefficients. American Psychologist, 45, 296.

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

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

Craft, N. P., & Kronenberger, E. J. (1979). Comparability of WISC-R and WAIS IQ with educable mentally handicapped adolescents. Psychology in the Schools, 16, 502 - 504.

Fine, M. A. (1992). On the distinction between two tests as measures of the same construct and as interchangeable. American Psychologist, 47, 1146.

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

Fuchs, D. & Fuchs, L. S. (1986). Test procedure bias: A meta-analysis of examiner familiarity effects. Review of Educational Research, 56, 243 - 262.

Gulliksen, H. (1950). Theory of mental tests. New York: John

Wiley.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Kangas, J., & 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.

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

Nagle, R. J., & Lazarus, S. C. (1979). II. The comparability of the WISC-R and WAIS among 16-year-old EMR children. The Journal of School Psychology, 17, 362 - 367.

Quereshi, M. Y., Treis, K. M., & Riebe, A. L. (1989). The equivalence of the WAIS-R and the WISC-R at age 16. Journal of Clinical Psychology, 45, 633 - 641.

Rubin, H. H., Goldman, J. J., & Rosenfeld, J. G. (1985). A comparison of WISC-R and WAIS-R IQs in a mentally retarded residential population. Psychology in the Schools, 22, 392 - 397.

Sandoval, J., Sassenrath, J., & Penaloza, M. (1988). Similarity of WISC-R and WAIS-R scores at age 16. Psychology in the Schools, 25, 373 - 379.

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

Sattler. J. M. (1992). Assessment of children: WISC-III and WPPSI-R supplement. San Diego: Sattler Publishing.

Sattler, J. M. & Gwynne, J. (1982). White examiners generally do not impede the intelligence test performance of black children: To debunk a myth. Journal of Consulting and Clinical Psychology, 50, 196 - 208.

Sattler, J. M. & Winget, B. M. (1970). Intelligence testing procedures as affected by expectancy and IQ. Journal of Clinical Psychology, 26, 446 - 448.

Slate, J. R. & Chick, D. (1989). WISC-R examiner errors: Cause for concern. Psychology in the Schools, 26, 78 - 84.

Slate, J. R & Hunnicutt, L. C. (1988). Examiner errors on the Wechsler scales. Journal of Psychoeducational Assessment, 6, 280 - 288.

Slate, J. R. & Jones, C. H. (1990). Student error in administering the WISC-R: Identifying problem areas. Measurement and Evaluation in Counselling and Development, 23, 137 - 140.

Slate, J. R., Jones, C. H., Coulter, C., & Covert, T. L. (1992). Practitioners' administration and scoring of the WISC-R: Evidence that we do err. Journal of School Psychology, 30, 77 - 82.

Slate, J. R., Jones, C. H., & Murray, R. A. (1991). Teaching administration and scoring of the Wechsler Adult Intelligence Scale - Revised: An empirical evaluation of practice administrations. Professional Psychology: Research

and Practice, 22, 375 - 379.

Tabachnick, B. G., & Fidell, L. S. (1989). Using multivariate statistics (2nd ed.). New York: Harper Collins.

Thompson, A. P., & Hodgins, C. (in press). Evaluation of a checking procedure for reducing clerical and computational errors on the WAIS-R. Canadian Journal of Behavioural Science.

Thompson, A. P., & Molly, K. (1993). The stability of WAIS-R IQ for 16-year-old students retested after 3 and 18 months. Journal of Clinical Psychology, 49, 891 - 898.

Votaw, D.F. (1948). Testing compound symmetry in a normal multivariate distribution. Annals of Mathematical Statistics, 19, 447 - 443.

Wechsler, D. (1955). The Wechsler Adult Intelligence Scale. Manual. New York: Psychological Corporation.

Wechsler, D. (1974). The Wechsler Intelligence Scale for Children - Revised. Manual. New York: Psychological Corporation.

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

Wechsler, D. (1991). The Wechsler Intelligence Scale for Children - Third Edition. Manual. New York: Psychological Corporation.

Appendix B

COMPARISON OF TWO INTELLIGENCE TESTS WITH 16 YEAR OLD STUDENTS

Dear Parent/Guardian:

I would like your permission to have your son/daughter participate in a research project. I am a Masters student in the Psychology Department at Lakehead University, an my supervisor is Dr. A.P. Thompson, Associate Professor and Registered Psychologist. The research is designed to determine whether the results of two intelligence tests are similar for 16 year old students. Your child would be asked to participate in two intelligence testing sessions over the next two (2) months. Each session lasts approximately 60 minutes. The tests will be administered individually, and in private.

Results of the tests will be confidential and will not be released to school officials. They are to be used solely for research purposes. We are interested in group trends rather than individual scores. It is unlikely that your son/daughter will find the intellectual assessment upsetting, although some people do feel anxious about their performance. Your son/daughter has the right to withdraw from the study at any time.

We will be willing to provide your son/daughter with a verbal explanation of their own results after they have completed

the second testing session. If they are interested in the overall results of the study, they may contact me. As well, they have the opportunity to win $150 in a draw for students participating in the study. Your son/daughter's participation in this study will reveal valuable information about intelligence testing which has not been thoroughly investigated.

If you approve of your son/daughter's participation in this research project, please complete the attached consent form. Your son/daughter can then return it to me. If you have any questions or concerns in relation to this research project, do not hesitate to contact me at 622-9962, or Dr. Thompson at 343-8646. Ethical approval for this research project has been received from the Lakehead University Ethics Committee and the Lakehead Board of Education.

Sincerely,

Daniela Sota

PARENT/GUARDIAN CONSENT FORM

I,_____, agree to allow my son/daughter

_____ to participate in the study on the use of intelligence tests with 16 year old individuals, conducted by Daniela Sota, Masters student, Lakehead University, and Dr. A.P. Thompson, Associate Professor, Lakehead University, and Registered Psychologist.

I understand that my son/daughter's participation in the study will involve being assessed intellectually on two separate occasions, each session running approximately 60 minutes.

SIGNATURE: _____

DATE: _____

COMPARISON OF TWO INTELLIGENCE TESTS WITH 16 YEAR OLD STUDENTS

Dear Participant:

Thank you for agreeing to participate in this research project. I am a Masters student in the Psychology Department at Lakehead University, and my supervisor is Dr. A.P. Thompson, Associate Professor and Registered Psychologist. The research is designed to determine whether the results of two commonly used intelligence tests are similar for 16 year old students. 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, lasting approximately 60 minutes, over the next two months. The tests are administered individually, and in private. The results of the tests will be confidential and will not be released to school officials. They are to be used for research purposes, and we are interested in group trends rather than individual scores. We are willing to provide you with a verbal explanation of your own results after you have completed the second testing session. Furthermore, if you are interested in the general results of this study, you may request a summary of the findings.

It is unlikely that you will find the intellectual

assessment upsetting, although some people do feel anxious about their performance. You have the right to withdraw from the study at any time. Participation in this research will give you first hand exposure to intelligence testing, which is a topic of much interest to psychologists. You will also have the opportunity to win $150 in a draw for students participating in the study.

Please sign the attached consent form, and have your parents sign the parental consent. If you have any questions or concerns, please feel free to contact me at 622-9962, or Dr. Thompson at 343-8646. Ethical approval for this research has been received from the Lakehead University Ethics Committee and the Lakehead School Board.

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

Sincerely,

Daniela Sota

PARTICIPANT CONSENT FORM

I, _____, agree to participate in the study on the use of intelligence tests with 16 year old individuals, conducted by Daniela Sota, Masters student, Lakehead University, and Dr. A.P. Thompson, Associate Professor, Lakehead University, and Registered Psychologist.

I understand that I will be assessed intellectually on two (2) separate occasions, each session running approximately 60 minutes.

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

SIGNATURE: _____

DATE: _____

HOME TELEPHONE: _____

Appendix C

I wanted to give you feedback on the first test that you took. First, I would like to define for you what intelligence is. Basically, it is your ability to solve problems. The intelligence test that you took breaks this ability down into two areas. The first is Verbal Intelligence, which is your ability to solve problems using words and numbers. The second is Performance Intelligence, which is your ability to solve problems, not using words, but visually, often by manipulating objects with your hands, like the puzzles you did.

Now your particular results were:
(feedback given only for first test, using 95% confidence intervals, and including a range only if the score fell into it by at least 3 IQ points)
1/ Your Verbal Intelligence was in the (score given according to the Wechsler classification) range.
2/ Your Performance, or non-verbal intelligence, was in the .. range.
3/ When you put these two results together, your overall problem-solving ability as measured by the test you took is in the .. range.

Do you have any questions? Would you like anything repeated?

This is not the absolute or final word on your intellectual level. There are four things that I'd like you to keep in mind. First, the conditions under which you took the test can influence the results, for example, if you were nervous, distracted or just not trying. Secondly, I would like to remind you that intelligence is only one factor related to success. You also need motivation and effort. So, people with high intelligence can squander their ability, and people with lower intelligence can be successful with persistence. Thirdly, the test you took doesn't measure all aspects of intelligence, such as musical aptitude, athletic ability, social skills, or aptitude to succeed in business. Lastly, you had just turned 16 when you took the first test, and we have done one other study which showed that 16 year olds who took a test like the one you took, usually increased when tested one and one-half years later. (The fourth point is only given if the student is 16 years, 3 months or younger.)

Just to make sure that you understand all that I just told you, I'd like you to repeat back to me what you learned.

If you have any more questions, I can be reached at 622-9962.